

Advances in Signal Processing and Artificial Intelligence

Proceedings of the 7th International Conference on Advances in Signal Processing and Artificial Intelligence (ASPAI' 2025)

Edited by Sergey Y. Yurish



Advances in Signal Processing and Artificial Intelligence:

Proceedings of the 7th International Conference on Advances in Signal Processing and Artificial Intelligence

> 8-10 April 2025 Innsbruck, Austria

Edited by Sergey Y. Yurish



Sergey Y. Yurish, *Editor* Advances in Signal Processing and Artificial Intelligence ASPAI' 2025 Conference Proceedings

Copyright © 2025 by International Frequency Sensor Association (IFSA) Publishing, S. L.

E-mail (for orders and customer service enquires): ifsa.books@sensorsportal.com

Visit our Home Page on http://www.sensorsportal.com

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (IFSA Publishing, S. L., Barcelona, Spain).

Neither the authors nor International Frequency Sensor Association Publishing accept any responsibility or liability for loss or damage occasioned to any person or property through using the material, instructions, methods or ideas contained herein, or acting or refraining from acting as a result of such use.

The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identifies as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

ISBN: 978-84-09-71189-5 BN-20250405-XX BIC: UYQ

Contents

Foreword
Compact Dual-band Millimeter Wave Antenna at Ka- and V-band for Sensing Applications7 <i>Parveez Shariff B. G., Tanweer Ali, Sameena Pathan and Pallavi R. Mane</i>
Kidney Tumor Segmentation Using Improved U-Net Architecture for Early Diagnosis of Renal Cell Carcinoma
Smart Sensor Selection: A Review on Metaheuristic Algorithms in IoT Platforms
Soft Computing for Flood Susceptibility Mapping of Kullu District of India19 Shweta Vincent, Mahesh Anil Inamdar, Om Prakash Kumar, Rohit Narayan H S, Nakul Rajendra Varma, Kaushik Naidu and Anadya Dang
Rényi Entropy-based Shrinkage Algorithm for Sparse Time-frequency Distribution Reconstruction Using Component Alignment Map24 V. Jurdana
Applied AI for DLT and CLT with Imperfect Bonding
Removing EOG Artifacts from EEG Recordings Using Deep Learning
GNSS Non-Line-of-Sight (NLOS) Error Repairing in Challenging Urban Environments with Channel Attention and Inception-based Deep Learning Network
Diagnosing Plant Leaf Disease with THz Sensor and Digital Signal Processing
Monitoring OoD Prediction Error in Semantic Segmentation Networks via Temporal Consistency of Logits
Examining Physiological Responses to Misophonic Triggers
Comparative Study of Route Algorithms Applied to Drones
CFUs Detection in Petri Dish Images Using YOLOv12
A Reliable and Efficient Detection Pipeline for Rodent Ultrasonic Vocalizations
Functional Connectivity Analysis Using Adaptive Window Size and Intersection of Confidence Intervals 79 Z. Šverko, S. Vlahinić, N. Stojković and P. Rogelj
Chart Pattern Recognition Using Convolutional Neural Networks
Prediction of Total Daily Diaper Changes Based on Infants' Bowel Sounds during the Beginning of Breastfeeding
Deep Jansen-Rit Parameter Inference for Model-driven Analysis of Brain Activity
Computing the Time-dependent Krankheit-operator in Epilepsy from ECoG: a Case Study

7th International Conference on Advances in Signal Processing and Artificial Intelligence (ASPAI' 2025), 8-10 April 2025, Innsbruck, Austria

Millimeter-wave Beam Prediction with Inverse Beamforming ML Model S. Mokdadi, S. E. Bouzid and P. Chargé	105
Video-based Analysis for Automated Ptosis Detection S. Baliński, P. Śniatała	111
Identification of Musical Instruments in Audios using Signal Analysis and Artificial Intelligence	115
Enhancing Real-time Decision-making with Scalable, Safe, and Private LLMOps and Context-aware RAG Workflows	119
Self-adaptive and Self-learning Lighting System: Integrating LSTM and RL for Energy Efficiency and Personalized Visual Comfort G. Potenza, Cristina Baglivo, M. Bonomolo and P. Ribino	125
Generation of a Rhythm Descriptor in Musical Phrases Using Signal Processing and Artificial Intelligence Techniques H. A. Aguilera-Garcia, R. A. Lizarraga-Morales	130
Combined Feature Selection and Hyperparameter Optimization for Small Datasets <i>N. L. Kämpf</i>	135
Res-Scrum: A Proactive and Resilient Agile Framework for Managing Uncertainty in Software Development <i>Aziz Fellah</i>	142
The Role of Code Readability in Large Language Model Code Summarization <i>B. Szalontai, G. Szalay, T. Márton, A. Sike, P. Mátray, M. I. Nagy, B. Pintér and T. Gregorics</i>	148
Traffic Predictions Using Graph Neural Networks on Real-time Observations Joachim Hansen, Donglin Liu and Alexandros Sopasakis	155
Knowledge Distillation for Efficient Algerian Dialect Processing: Training Compact BERT Models with DziriBERT Laggoun Amina, Zakaria Chahnez and Smaili Kamel	161
An Evaluation of General-purpose Large Language Models for Aspect Summarization S. Frank, C. Gütl and A. Wagner	167
Characteristics of Dynamic Velocity Response in Hand Movements Using Frequency and Time Modeling Techniques C. L. Sandoval-Rodriguez, A. F. Jimenez-Quezada, N. Orejarena-Osorio, O. Lengerke, and D. M. Rey Bravo	171 es-
Graphical User Interface for Volumetric Capnography: Parameter Estimation and Fowler's Method Implementation C. L. Sandoval-Rodriguez, N. Orejarena-Osorio, A. F. Jimenez-Quezada, and O. Lengerke	176
Forecasting Flood in Vietnam Using Deep Learning T. L. Nguyen, T.H. Nguyen	180
Enhancing Accuracy in Non-contact Physiological Monitoring: The Critical Role of Radar and Sensor Signal Alignment Nour Ghadban, Mostafa Elsayed, Jonathan Cooper, and Julien Le Kernec	184
Radial Basis Operator Networks J. A. Kurz, S. Oughton and S. Liu	189
The Protocol for Integration of Automated and Dynamic Facial Expression Emotion Recognition with EEG for Emotional Traits Analysis in Pilot Candidates S. Michalak, T. Łodygowski, P. Śniatała, M. Goralewski, E. Kozielewska-Zwierska, J. Moskal, M. Galant-Gołebiewska, M. Maciejewska. K. Śniatała, P. Zvch	197
Neurorehabilitation System Supported by Virtual Reality P. Śniatała, S. Michalak, E. Kozielewska-Zwierska, A. Krawczyński, K. Śniatała, S. Baliński	201
Radioactive Tabular Datasets to Detect Unauthorized Machine Learning Mehdi Ben Ghali, Gouenou Coatrieux and Reda Bellafqira	206

MineralBLIP: Advancing Mineral Classification with Vision Language Pre-training Model21	2
Khalid Alharthi [,] Ghadi Alkhushail, Sharifah Malhan, Batol Alsalkhadi, Hatun Alqarni,	
Kholoud Alharthi, Reem Almarhabi, Raghad Alharthi, Ali Alshahrani, Muhammad Zaka Emad, and Dhafer Alshehri	

An Improved Algorithm for Computing Matroids over Polynomials	218
David W. Ash	

Foreword

It is with great pleasure, enthusiasm and pride that I present the proceedings of the 7th International Conference on Advances in Signal Processing and Artificial Intelligence (ASPAI' 2025), held in the beautiful city of Innsbruck, Austria, from April 8 to 10, 2025.

The ASPAI Conference Series has become a significant international forum for the dissemination and exchange of cutting-edge research in the domains of signal processing and artificial intelligence—two pillars of modern information and decision systems. These fields continue to evolve synergistically, offering transformative capabilities across critical sectors such as biomedical diagnostics, remote sensing, cognitive robotics, telecommunications, environmental modeling, and autonomous navigation systems.

The proceedings compiled herein reflect the diversity and technical depth of contemporary research at the intersection of these domains. Each paper has been rigorously peer-reviewed and selected for its originality, scientific merit, and potential impact. Contributions span from advanced time-frequency analysis and millimeter-wave antenna design to innovations in deep neural architectures, federated learning systems, and intelligent optimization for Internet of Things (IoT) platforms. These works not only demonstrate theoretical advancements but also emphasize real-world applicability and scalability—highlighting the robust interplay between foundational theory and applied engineering.

In an era where data volumes are exponentially increasing and computational intelligence must be both adaptive and interpretable, the work presented at ASPAI' 2025 offers substantive insights into the design of resilient, efficient, and ethical AI-driven systems. Particular attention has been given to frameworks that enhance model robustness, privacy preservation, energy efficiency, and real-time signal interpretation—underscoring the community's dedication to solving complex challenges under realistic constraints.

The sustained excellence and impact of ASPAI are made possible through the committed efforts of our authors, the meticulous work of our reviewers, and the strategic vision of the conference organizing committee. I express my sincere gratitude to all contributors and collaborators who have supported this endeavor with their time, expertise, and academic rigor.

As the Chairman of ASPAI' 2025, I am confident that this volume will serve not only as a reference for current state-of-the-art methodologies but also as a catalyst for new ideas, collaborations and novel research trajectories. I encourage readers to engage deeply with the material and to explore interdisciplinary synergies that drive innovation at the nexus of signal processing and artificial intelligence.

I extend my sincere thanks to all the authors, reviewers, organizing committee members, and sponsors who have contributed to the success of this conference. Your dedication and hard work are deeply appreciated.

Prof., Dr. Sergey Y. Yurish ASPAI' 2025 Conference Chairman (006)

Compact Dual-band Millimeter Wave Antenna at Ka- and V-band for Sensing Applications

Parveez Shariff B. G.¹, <u>Tanweer Ali</u>¹, Sameena Pathan² and Pallavi R. Mane¹

 ¹ Department of Electronics and Communication Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal 576104, India
 ² Department of Information and Communication Technology, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal 576104, India Tel.: +91-8050362729 E-mail: tanweer.ali@manipal.edu

Summary: With technological advancement, devices are becoming intelligent with many sensors onboard. The sensors push Gbps of data on the internet every day. The current sub-6 GHz frequency band has reached saturation due to bandwidth constraints. As a result, a millimeter wave (mmWave) spectrum with licensed and unlicensed bands is exposed to various applications. Thus, the article presents a compact with a small form factor of $0.66\lambda_1 \times 0.64\lambda_1$, having dual-band resonance and operating in Ka and V-band. The antenna achieved a fractional bandwidth of 200 % at both bands with a maximum gain of 6 dBi. Thus, the proposed antenna is suitable for sensing applications due to its compact structure at the mmWave spectrum.

Keywords: 5G, Ka-band, Millimeter wave antenna, V-band.

1. Introduction

The number of smart devices around us is rapidly increasing and is expected to reach 25.44 billion by 2030. These devices demand continuous connectivity to the external world, pushing out Gbps of data every day. Most of these devices are connected to multimedia, vehicular communication (V2V and V2X), payment terminals, tracking and monitoring stations, inventory management devices, etc. The International Telecommunication Union (ITU) has opened the higher millimeter wave (mmWave) spectrum from Ka- to E-band to accommodate substantial data requirements [1], increasing the bandwidth to 10-fold.

The mmWave spectrum is primarily studied for communication; however, its potential is later explored in sensing applications, such as on-body health monitoring, driver alertness, in-door movement monitoring, an inspection of building cracks, etc. [2]. The transducer for sensing applications must be compact to embed in compact devices. Thus, an antenna characterized as transducer is presented in this article. For example, in [3], a compact dual-band antenna for on-body application is proposed. The first band is 34 GHz, and the second is 60 GHz. However, it resulted in a narrow bandwidth. In [4], the gain of an antenna at 60 GHz is increased by arranging the radiating elements in an array fashion. The design adopts the substrate-integrated waveguide (SIW) feed mechanism, making it complex in fabrication. Further, to simplify the complexity, in [5], a planar antenna is designed to resonate at three bands, possessing decent bandwidth at least in two bands. The design adopted a co-planar waveguide structure for design simplicity.

2. Antenna Design

In contrast to the above literature, this article presents a planar compact dual-band antenna operating in Ka and V-band, as shown in Fig. 1, for mmWave sensing applications.



Fig. 1. Proposed antenna design with top view in (a) and bottom view in (b).

The internal dimension of the antenna is presented in Table 1. The design is etched on Roggers 5880 substrate, which has a thickness of 0.254 mm. The antenna profile is $0.66\lambda_1 \times 0.64\lambda_1$ (where λ_1 is wavelength at 38 GHz). The radiator has two elliptical rings interconnected by vertical stubs. From the internal ring, a microstrip line connects the two openended monopole-like structures responsible for generating dual resonance. However, the bandwidth at these bands was narrow; as a result, the ground plane is defected to improve the bandwidth. The resulting reflection coefficient is illustrated in Fig. 2.

Table 1. Antenna dimensions in mm.

Parameter	Value	Parameter	Value	Parameter	Value
FL	0.95	FW	0.25	E1L	4
E1H	1.6	E2L	2.72	E2H	1.31
E3L	2.1	E3H	1	E4L	1.82
E4H	0.6	T1	1.25	D1	2
D2	2.74	S1	1.4	S2	1
S3	2.5	S4	.4	S5	1.4
S6	1.85	S7	1.2	S8	1



Fig. 2. Simulated reflection coefficient of the proposed antenna.

The designed antenna achieved a bandwidth range of 34.3-42.5 GHz and 49.4-53.3 GHz. Thus, at both bands, the fractional bandwidth is 200 %. The antenna has bi-directional radiation characteristics in the E-plane and H-plane at 38 and 52 GHz, with a maximum gain of 5.2 and 6 dBi, respectively, as shown in Fig. 3.

3. Conclusion

The article presented a compact $0.66\lambda_1 \times 0.64\lambda_1$ antenna with dual-band resonance in Ka and V-band with 200 % fractional bandwidth. The antenna has gain of 6 dBi and suitable for mmWave sensing applications.



Fig. 3. Simulated radiation characteristics at 38 and 52 GHz in (a) E-plane and (b) H-plane.

References

- N. K. Mallat, M. Ishtiaq, A. Ur Rehman A. Iqbal, Millimeter-wave in the face of 5G communication potential applications, *IETE Journal of Research*, Vol. 68, Issue 4, Jul. 2022, pp. 2522-2530.
- [2]. B. Van Berlo, A. Elkelany, T. Ozcelebi, N. Meratnia, Millimeter wave sensing: a review of application pipelines and building blocks, *IEEE Sensors J.*, Vol. 21, Issue 9, May 2021, pp. 10332-10368.
- [3]. U. Farooq, G. M. Rather, A miniaturised Ka/V dual band millimeter wave antenna for 5G body centric network applications, *Alexandria Engineering Journal*, Vol. 61, Issue 10, Oct. 2022, pp. 8089-8096.
- [4]. J. Hautcoeur, A. Ghayekhloo, K. Hettak, L. Talbi, H. Boutayeb, K. Wu, 60 GHz frequency sensor antenna for short-range millimeter-wave detection application, *IEEE Sens. Lett.*, Vol. 6, Issue 10, Oct. 2022, pp. 1-4.
- [5]. S. Ahmad, *et al.*, Design of a tri-band wearable antenna for millimeter-wave 5G applications, *Sensors*, Vol. 22, Issue 20, Oct. 2022, 8012.

(007)

Kidney Tumor Segmentation Using Improved U-Net Architecture for Early Diagnosis of Renal Cell Carcinoma

Sameena Pathan¹, Tanweer Ali² and Haneena Hyder¹

 ¹ Department of Information and Communication Technology, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal 576104, India
 ² Department of Electronics and Communication Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal 576104, India Tel.: +91-9972208602 E-mail: sameena.bp@manipal.edu

Summary: The Kidneys are vital organs that remove waste products and excess fluid from the circulation, which is crucial for preserving health. Renal Cell Carcinoma (RCC), sometimes known as kidney cancer, is the most frequent type of adult cancer, accounting for 3-4 % of cases. Particularly in males over 64, a large number of cases are asymptomatic and inadvertently discovered. Smoking, obesity, and a bad diet are risk factors, and the chances of survival differ greatly depending on the stage. The most important diagnostic methods are CT and MRI, and early detection is essential. Techniques for segmenting images improve analysis by concentrating on particular regions. The accuracy of kidney tumor segmentation and diagnosis has improved recently due to developments in deep learning and automated analysis, particularly with Convolutional Neural Networks (CNNs). This aids in the more precise pathology diagnosis made by doctors.

Keywords: Kidney tumor segmentation, Deep learning, U-Net, KiTS19 dataset, Medical imaging.

1. Introduction

Kidneys are essential organs responsible for filtering waste products and excess fluids from the blood, playing a critical role in regulating electrolyte balance, blood pressure, and red blood cell production. However, they are also vulnerable to diseases such as kidney cancer, particularly Renal Cell Carcinoma (RCC), which accounts for approximately 3-4 % of adult cancers. RCC is notably the third most common urological cancer, with clear cell RCC being the most prevalent subtype, making up 80-90 % of cases. This type of cancer often develops with minimal symptoms, leading to many cases being detected incidentally, particularly in men over the age of 64. Key risk factors include smoking, obesity, poor dietary habits, and a family history of hypertension.

Given the challenges associated with manual tumor segmentation, which is time-consuming and prone to subjectivity, this research underscores the urgent need for automated segmentation techniques to enhance the accuracy and efficiency of kidney tumor identification. The study aims to leverage advancements in deep learning and image analysis to develop robust algorithms capable of accurately segmenting kidney tumors from CT scans, followed by a size-based analysis for tumor staging. By comparing various segmentation algorithms, the research will evaluate their performance across different contexts and establish a standardized approach for automated classification.



Fig. 1. CT scan of both kidneys and tumour.

2. Literature Survey

Kidney segmentation techniques in medical imaging, particularly using CT scans, have evolved significantly, incorporating classical image processing methods as well as advanced deep learning techniques. Kaur and Juneja [1] conducted an extensive survey reviewing various kidney segmentation methods, discussing both traditional and modern approaches, and identifying their advantages and limitations, such as dependency on manual intervention and sensitivity to image quality. Chow et al. [2] contributed to the broader understanding of renal cell carcinoma by identifying epidemiological risk factors, facilitating early detection and prevention strategies. Pan et al. [3] identified miR-566 as a molecular biomarker, highlighting its potential role as an oncogene and prognostic indicator in renal cell carcinoma.

Further advancements in deep learning techniques have substantially enhanced medical image segmentation accuracy and efficiency. Müller and Kramer [4] introduced MIScnn, a dedicated framework for efficient and precise segmentation using convolutional neural networks (CNNs). Hesamian et al. [5] reviewed the substantial of CNNs in achievements medical image segmentation, emphasizing the associated challenges such as computational intensity and dataset requirements. Zhu et al. [6] successfully demonstrated improvements in renal tumor segmentation accuracy using transfer learning with CNNs, particularly beneficial for small dataset scenarios. The KiTS19 challenge dataset by Heller et al. [7] provided standardized, clinically annotated datasets, significantly propelling research in kidney tumor segmentation. Bolocan et al. [8] further advanced segmentation techniques by presenting a CNN-based model specifically designed for effective segmentation and classification of clear cell renal cell carcinoma using multiphase CT images. Alzu'bi et al. [9] introduced a new dataset focusing specifically on kidney tumor detection and classification using deep learning, expanding resources available for research. Additionally, Pandey and Gupta [10] demonstrated an effective approach to tumorous kidney segmentation using active contour methods integrated with 3D-UNet models, further advancing segmentation capabilities in medical imaging.

2. Methodology

The methodology uses techniques such as random cropping for dataset augmentation, resolution filtering, and intensity value scaling when processing CT scans. Techniques for augmenting data, such as rotations and flips, raise variability. Tumor size analysis based on the TN staging method comes after segmentation using deep learning models like U-Net and V-Net.

2.1. Dataset Description

The dataset consists of multi-phase CT imaging, i.e., it includes CT scans captured with different contrast agents, offering detailed information on various types of tissues. Each patient case includes a CT scan and the respective segmentation masks. These masks are essentially digital labels that precisely outline the regions of interest in the CT images. In this case, the masks likely depict both the kidney contours and the tumor boundaries. The dataset also includes comprehensive clinical outcomes for the 210 patients. The CT scans and segmentation masks are provided in the anonymized NIFTI format. This is a common file format in medical imaging tasks as it is suitable for storing volumetric data like 3D medical images. The NIFTI format typically uses a shape representation of (num slices, height, width) [7]. This indicates that the data is organized as a 3D volume, with each entry representing:

- num_slices: The total number of individual CT slices in the scan;
- height: The number of pixels in the vertical direction (rows) of each slice;
- width: The number of pixels in the horizontal direction (columns) of each slice.

2.2. Data Processing

The dataset underwent formatting and cleaning, including size-based inclusion step to filter data points. Min-Max scaling was applied for standardization, and data augmentation techniques were used to enhance the dataset. This workflow is depicted in Fig. 2. The model finds it challenging to learn efficiently as a result. The CT images intensity values were scaled to a specific range (between -57 and 164 based on the code) for better model convergence during training using min-max scaling method, where x is the intensity value, x.min() is the minimum intensity value, and x' is the scaled intensity value as given in (1)

```
x' = (x - x.\min()) \div (x.\max() - x.\min()) (1)
```



Fig. 2. Flowchart describing the sequential process in the study.

Random Cropping: This technique helps to an increase the dataset size and improve model generalizability by exposing it to various image excerpts.

2.3. Data Augmentation

The code introduced random flips (along horizontal, vertical, and depth axes) and rotations (up to 90 degrees) during data augmentation. Random Flips is the process of flipping the image along horizontal, vertical, and depth axes essentially creates mirrored versions of the original image. This exposes the model to how the kidney and tumor might appear if viewed from a different perspective. Random Rotations rotates the image by up to 90 degrees simulates potential variations in how the patient was positioned during the CT scan acquisition. This helps the model learn features that are rotation-invariant, allowing it to accurately segment the kidney and tumor regardless of their in-plane orientation within the image.

2.4. Segmentation

The processed data was split into training, validation, and testing sets. Various segmentation

7th International Conference on Advances in Signal Processing and Artificial Intelligence (ASPAI' 2025), 8-10 April 2025, Innsbruck, Austria

algorithms, such as U-Net, were applied to accurately delineate tumor boundaries.

2.4.1. U-Net

U-Net is widely known for its effectiveness in medical segmentation. Its strength lies in the skip connections that transfer spatial information directly from the encoder to the decoder as given in Fig. 3 and 4. The performance of the U-Net across 100 training epochs, measured by Dice loss and IOU score.





Fig. 4. Encoder Architecture.

2.4.2. V-Net

V-Net is designed specifically for 3D medical image segmentation and builds upon U-Net's structure with some key enhancements. The performance of the V-Net across 100 training epochs, measured by Dice loss and IOU score.

2.4.3. Attention U-Net

Attention U-Net adds an attention mechanism to the traditional U-Net to help the model focus on more relevant regions within the image.

2.4.4. Autoencoder

Autoencoders can also be adapted for segmentation by learning compact representations of the input and reconstructing the segmentation mask from these features.

2.4.5. Shared Design Principles

While each model has its own architectural nuances, they follow a similar high-level process:

- 1. Downsample the input to extract deep features (Encoder);
- 2. Process or refine those features using dense layers, attention, or bottlenecks;
- 3. Upsample and combine with earlier features for detailed reconstruction (Decoder);
- Generate a segmentation mask through a final output layer.

3. Results

3.1. Performance Metrics

In our research investigation, we employed the segmentation algorithms U-Net, V-Net, Attention U-Net, and Auto Encoder. The CT scan patch image, the label (ground truth), and the tumor class prediction are all included in the outcome. It evaluates the degree to which the model has correctly recognized the tumor by contrasting the prediction with the label (ground truth). The tumor pixels are then identified using the matching label and prediction masks, and the size of the label tumor and the predicted tumor are computed by iterating through each patch. The tumor's overall size in each patch is then calculated by adding up all of the tumor's pixels, and the TNM staging system is used to categorize the tumor into various stages.

The performance of the segmentation models was evaluated using established metrics, as shown in Table 1. The analysis provided insights into the strengths and weakness of each algorithm, guiding future improvements.

Table 1. Results of segmentation algorithms.

Segmentation algorithm	Dice Score	Dice Loss	IOU Score
Auto Encoder	0.82	0.17	0.69
V-Net	0.96	0.03	0.92
U-Net	0.97	0.02	0.95
Attention U-Net	0.98	0.01	0.96

The strongest model, Attention U-Net, obtained an IOU score of 0.9654 and a Dice score of 0.9824 when evaluated on unseen test data. The tumor prediction made by the model based on the test data is shown in Fig. 5.

3.2. Result Analysis

From the above dice scores and IOU score we can compare and obtain that the Attention U-Net model performs best on the dataset.

After the predictions were obtained, the tumor size of the label was measured along with the tumor size of the predicted model and compared and put into different stages according to the TNM classification.

3. Conclusion

The study effectively illustrates the use of deep learning methods for comparative size analysis and automated kidney tumor segmentation in kidney cancer staging. Tumor boundaries are essential for proper diagnosis and treatment planning. The research significantly improved tumor boundary delineation by utilizing the KiTS19 dataset and sophisticated segmentation algorithms such as U-Net and Attention U-Net. Subsequent research endeavors mav concentrate on optimizing these models, investigating supplementary data sources, and including clinical characteristics to augment the resilience and suitability the segmentation techniques of in actual medical contexts.





References

- R. Kaur, M. Juneja, A survey of kidney segmentation techniques in CT Images, *Curr. Med. Imaging Rev.*, Vol. 14, 2017, pp. 238-250.
- [2]. W. H. Chow, L. Dong, S. Devesa, Epidemiology and risk factors for kidney cancer, *Nat. Rev. Urol.*, Vol. 7, 2010, pp. 245-257.
- [3]. X. Pan, J. Quan, Z. Li, et al., MiR-566 functions as an oncogene and a potential biomarker for prognosis in renal cell carcinoma, *Biomed. Pharmacother.*, Vol. 102, 2018, pp. 718-727.
- [4]. D. Müller, F. Kramer, MIScnn: a framework for medical image segmentation with convolutionaneural networks and deep learning, *BMC Medical Imaging*, Vol. 21, 2021, 12.
- [5]. M. H. Hesamian, W. Jia, X. He, P. Kennedy, Deep learning techniques for medical image segmentation: achievements and challenges, *J. Digit. Imaging.* Vol. 32, 2019 pp. 582-596.
- [6]. X.-L. Zhu, H.-B. Shen, H. Sun, et al., Improving segmentation and classification of renal tumors in small sample 3D CT images using transfer learning with convolutional neural networks, *International Journal* of Computer Assisted Radiology and Surgery, Vol. 17, Issue 7, 2022, pp. 1303-1311.
- [7]. N. Heller, N. J. Sathianathen, A. A. Kalapara, et al., The KiTS19 challenge data: 300 kidney tumor cases with clinical context, CT semantic segmentations, and surgical outcomes, *arXiv preprint*, 2019, abs/1904.00445.
- [8]. V.-O. Bolocan, M. Secareanu, E. Sava, et al., Convolutional neural network model for segmentation and classification of clear cell renal cell carcinoma based on multiphase CT images, *J. Imaging*, Vol. 9, 2023, 280.
- [9]. D. Alzu'bi, M. Abdullah, I. Hmeidi, et al., Kidney tumor detection and classification based on deep learning approaches: a new dataset in CT scans, *Journal of Healthcare Engineering*, Vol. 2022, 2022, 3861161.
- [10]. M. Pandey, A. Gupta, Tumorous kidney segmentation in abdominal CT images using active contour and 3D-Unet, *Ir. J. Med. Sci.*, Vol. 192, Issue 3, 2023, pp. 1401-1409.

(010)

Smart Sensor Selection: A Review on Metaheuristic Algorithms in IoT Platforms

Sujith Kumar¹, Shweta Vincent¹ and Om Prakash Kumar²

 ¹ Department of Mechatronics, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, 576104, Karnataka, India
 ² Department of Electronics and Communication Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, 576104, Karnataka, India

E-mail: shweta.vincent@manipal.edu

Summary: Despite tremendous advances, existing techniques to sensor selection in Internet of Things (IoT) systems confront numerous obstacles. Multi-objective evolutionary algorithms, such as MOEA/D and NSGA-III, excel in solving complicated problems with numerous objectives. Evolutionary algorithms improve solution quality, but they have limitations due to parameter biases and constrained applicability. Improving energy optimization algorithms can increase resource efficiency and network lifespan, but scaling and executing them in real-world circumstances presents considerable obstacles. Sampling and visualization approaches provide useful insights, albeit with limitations such as data loss and the presumption of uniform solution distribution.

Keywords: Multi-objective optimization, Evolutionary algorithms, Design of algorithm, Analysis, Performance evaluation.

1. Introduction

Multi-objective evolutionary algorithms, or MOEAs, are a key feature of evolutionary computing that has piqued the interest of both scholars and practitioners.

1.1. Introduction to Multi-objective Optimization

Over the last two decades, tremendous progress has been made in developing algorithms capable of handling optimization. Difficulties can arise during complex decision-making processes in a variety of fields, including engineering, finance, healthcare, and others.

1.2. Domains of Multi-objective Optimization

Decision-makers must strike a balance between competing demands such as resource allocation, system efficacy, and cost efficiency in some sectors where critical decision making is required. Multi-Objective Evolutionary Algorithms (MOEAs) seek to address these difficulties by combining computational intelligence and natural evolution principles to provide cost-effective and scalable solutions that achieve nearly optimal trade-offs between many objectives. They are widely used in a variety of industries, including control systems, renewable energy, process optimization, and structural design. Portfolio management demonstrates multiobjective optimization in finance, with investors seeking to maximize returns while limiting risk. Similarly, healthcare applications such as developing treatment planning and allocating resources necessitate

striking a balance between patient outcomes, resource utilization, and cost effectiveness.

In addition to these applications, algorithms are widely used in machine learning, environmental management, transportation planning, and supply chain optimization. Their adaptability is critical for resolving the ever-increasing complex and interconnected difficulties confronting modern decision-makers.

1.3. Sensor Placement in IoT Using Multi-objective Optimization

The complexities of real-world scenarios highlight the importance of efficient optimization approaches in the current context. One of the primary issues in the world of the Internet of Things (IoT) is the accurate location of sensors in order to improve coverage, speed data collecting, and ensure the system's smooth operation. The primary goal is typically to ensure that the region or environment where IoT data must be collected is sufficiently covered such as temperature, humidity, and motion, as well as their geographical and temporal precision, as well as any limitations on data quality and dependability. The physical environment is one of the elements that influence sensor positioning. Sensor placement must be carefully considered in order to reduce signal disruption and provide dependable network communication. It may be important to install battery-powered sensors in locations where they can be quickly replaced or recharged. Automatic optimization algorithms, such as heuristic and metaheuristic techniques, can assist in automating and improving sensor site selection. These algorithms use characteristics such as coverage, connectivity, and energy efficiency to find the most ideal or near-optimal solutions. Modeling and simulation techniques are effective for forecasting the efficacy of various sensor placements before they are implemented.

1.4. Motive

1. Precision of Data: Sensors are strategically placed to ensure a continual stream of reliable data. In environmental monitoring, correct sensor placement is critical for reliably assessing contaminant levels.

2. Cost effectiveness: Because of the large number of sensors used in IoT systems, cost-effectiveness is frequently a critical consideration. By carefully selecting appropriate places, it is feasible to reduce the number of sensors required.

3. Energy Efficiency: Many IoT devices rely on low-power sources or batteries, thus energy economy is an important consideration. Inadequately located sensors may cause faster battery drain resulting in higher operational costs and increased maintenance requirements.

4. Network capacity: The network capacity required to transfer data from sensors to the central processor is critical. Sensors can be strategically placed to reduce superfluous data transfers, alleviate network congestion in large-scale IoT systems, and ensure that only critical data is transferred.

5. Robustness and Reliability: The IoT system's robustness and dependability improve when strategically placed sensors are used. They outperform other elements in the environment in terms of their ability to tolerate difficult situations, interruptions, and impediments. When sensors are properly located, the likelihood of sensor malfunction and data loss due to external influences is reduced.

6. Security and Privacy Concerns: Incorrectly configured sensors have the possibility to accidently collect personal information or violate privacy standards. Strategic sensor placement is critical for reducing the possibility of breaches and ensuring compliance with security and privacy requirements.

7. Scalability and Flexibility: Strategic sensor placement increases the scalability and flexibility of IoT systems. A carefully designed placement strategy allows for smooth development or adjustment without requiring extensive redesign as the system evolves or its requirements change.

Finally, carefully selecting sensor placements in Internet of Things (IoT) systems is crucial

1.5. Methodology Implementation

This review employs a novel approach by carefully evaluating and contrasting several multi-objective and multi-objective evolutionary algorithms. The purpose is to provide a thorough understanding of sophisticated methodology and their applications to complex optimization problems by organizing the analysis around key concepts, approaches, and empirical findings. The expected conclusions of this review will be immensely beneficial to scholars, practitioners, and decision-makers interested in multi-objective optimization. The review's purpose is to foster information interchange, encourage multidisciplinary teamwork, and ignite innovative approaches to addressing complicated optimization issues across disciplines by incorporating insights from many sources.

2. Problem Definition

This study focuses on solving key issues and gaps in the field of multi-objective algorithms for sensor selection in IoT systems. The difficulties originate from the complex nature while optimizing systems. By addressing these challenges, the initiative hopes to demonstrate the significance and relevance of its objectives.

Traditional single-objective optimization approaches frequently fail when dealing with complex real-world circumstances. The research focuses on a fundamental issue such as the inadequate assessment of the strengths and limits of current algorithms for sensor selection in IoT applications. Although many algorithms have been created over the years, without an extensive study, the actual application and improvement of these algorithms are restricted.

3. Objectives

The study's primary purpose is to create, analyze, implement, validate, and disseminate improved sensor selection optimization methods for IoT systems. Examine the existing multi-objective design methodologies for choosing IoT sensors, to determine how effectively the algorithm performs and record the findings in a review article to share.

4. Theoretical Background

4.1. Evolutionary Algorithm for Multiobjective Optimization

Evolutionary algorithms (EAs) are optimization methods inspired by natural evolution. They iteratively create new solutions by applying selection, iteration, and transformation operators to a pool of candidate solutions. Multi-objective and multi-objective optimization problems necessitate the simultaneous optimization of several conflicting objectives, making traditional single-objective optimization methodologies inefficient. Evolutionary algorithms, particularly multi-objective evolution algorithms (MOEA) and multi-objective evolution algorithms (MaOEA), have developed as viable strategies for dealing with such complex optimization problems.

4.1.1. Multi-objective Evolutionary Algorithms (MOEA)

The goal of MOEAs is to identify a set of Pareto-optimal solutions that reflect the trade-offs between competing goals. Popular MOEAs include the Non-Dominated Sorting Genetic Algorithm-II (NSGA-II), Strength Pareto Evolutionary Algorithm (SPEA), and Decomposition-Based Multi-Objective Evolutionary Algorithm. These algorithms preserve a varied range of non-dominated solutions, encouraging research and application of the Pareto frontier.

4.1.2. Many-objective Evolutionary Algorithms (MaOEA)

Multi-Objective Evolutionary Algorithms (MOEAs) enable the solution of optimization problems involving several objectives, often more than three. Several solutions have been proposed to address multi-objective optimization difficulties, including NSGA-III, MOEA/D with increased decomposition, and indicator-based evolutionary algorithms. Multi-Objective Evolutionary Algorithms (MOEAs) seek to support a wide range of decentralized solutions throughout the whole Pareto front, assuring the most efficient use of computational resources.

4.1.3. Opposition Based Learning (OBL)

Opposition-based learning (OBL) is a heuristic problem-solving approach inspired by the concept of oppositions. The Optimization by Learning approach pairs each solution with its complementary counterpart within the search space, allowing for a comparison of their performances to determine which is best. OBL is used in evolutionary algorithms to increase the quality of solutions, expedite convergence, and preserve diversity.

4.2. Mathematical Formulation

The simplest and most generic equation utilized in the optimization techniques is as follows.

Minimize the solution in a region (R),

$$F(x) = (f1(x), f2(x), \dots, fn(x),$$
(1)

where x are the variable vectors and $x \in R$, F(x) is the Objective function, fi is the ith objective, i = 1,2,3,...n, is the Region defined in the problem.

Typically, the theoretical framework provides a solid foundation for understanding the essential concepts, techniques, and mathematical models of multi- and multi-objective optimization. These notions serve as the foundation for the creation and testing of optimization strategies for sensor selection in IoT systems.

5. Methodology

5.1. Literature Review Methodology

The approach used in this research involved a comprehensive and detailed review.

- The approach used is described in stages:
- 1. To identify base and supplemental papers: An extensive search was conducted on academic databases such as IEEE Xplore, ACM Digital Library, ScienceDirect, and Google Scholar using relevant keywords;
- 2. Data extraction and synthesis: Obtaining relevant material from selected papers, including algorithm descriptions and main conclusions;
- 3. Assessment of Quality: The publications included in the review were evaluated for quality in terms of trial design, methodological clarity, result interpretation, and citation impact.

5.2. Result Analysis

The preliminary data analysis provides a comprehensive summary of the key findings and trends identified in the selected literature. This work adds value by finding parallels, contrasts, and potential for improvement in current multi- and multi-objective optimization methods, allowing for more in-depth analysis and debate in the next sections of the review paper.

6. Literature Review

Lin et al. 2017, presented a multi-objective technique for optimizing sensor selection in IoT systems. The study examined a range of multi-objective algorithms on diverse datasets with varying sensor and component counts. Despite considerable progress, the proposed approach is still constrained by scale issues. The literature analysis indicates the study's shortcomings, particularly in information coverage, energy consumption, and network lifetime. The examination of IoT sensor selection methods across different datasets found a mix of convergence and variation, such as hypervolume [1].

The data offered contains both actual sensor data and carefully designed produced data. They have applications in a variety of disciplines, including environmental monitoring and control. While many utility-based sensor selection approaches are NP-hard, optimizing certain features can be done effectively. Although this strategy produces positive results, it is restricted by limitations such as high processing costs, a limited utility function. More study is needed to address these issues, and researching optimization methodologies. Improvements in these areas will increase the practicality and efficiency of picking sensors based on utility [2].

Sensor selection approaches improve sensor performance by combining binary and continuous data into a chromosome for optimization. The efficiency of these strategies is evaluated in simulated environments using datasets that include sensor features, network structure details, and varied spatial layouts, as well as changing densities and ambient elements. Genetic algorithms using a combination of genetic material are used to improve a set of sensors in wireless sensor networks by taking into account their fitness ratings and the number of nodes. Nonetheless, dealing with varied node designs and population dynamics offers new issues that must be investigated and proven in real-world circumstances. Using various selection procedures has the potential to increase the adaptability of sensor choosing algorithms [3].

Recent breakthroughs in multi-objective evolutionary algorithms have led to the development of MOEA/D. This method entails splitting down issues into scalar optimization subproblems. To address scalability and diverse scaling goals, many decomposition methods are being studied, including weighted sum. MOEA/D surpasses other algorithms, such as MOGLS and NSGA-II, in terms of computing complexity and solution quality, yielding equally dispersed Pareto optimum solutions. Furthermore, more research is needed to acquire a thorough understanding of the scalability and sensitivity in respect to neighborhood size [4].

The NSGA-III algorithm uses a benchmark technique, has recently gained attention as a feasible option. When it comes to convergence and diversity, NSGA-III regularly outperforms MOEA/D algorithms on DTLZ test problems with diverse aims. The study could benefit from more in-depth evaluations when compared to other complex algorithms. Despite these constraints, NSGA-III excels at tackling optimization problems many objectives while ensuring variety and convergence across several test cases. More research is needed to establish its applicability in real-world [5].

The paper emphasizes the versatility and efficacy of multi-population techniques in tackling real-world optimization difficulties such as scheduling, path planning, network optimization, and parameter estimation. The work addresses persistent issues such as scalability, processing efficiency, parameter sensitivity, and convergence. It is critical to address these limitations in optimization and design applications [6].

The authors of [7], created a mechanism for mapping energy-sensitive locations and selecting sensors. This strategy intends to improve sensor deployment by assigning sensors for energy efficiency and dynamically modifying task duration to reduce energy consumption. While the method has not been confirmed by simulations, it does show promise for improving energy efficiency. Despite the issues, author emphasizes the importance of future research and development efforts. Yu et al. (2014) offer a Quality of Service (QoS)-based technique for improving sensor selection. They consider sensor durability, battery utilization, and data transmission quality. The researchers compared the efficiency of an ILP-based matching service technique and a greedy matching algorithm for IoT devices that used synthetic materials. The findings showed that, while ILP improved QoS adaption, it required more computing work, resulting in the development of more efficient greedy algorithms. The study's findings highlight the difficult balance between solution efficacy and computational efficiency in the integration of IoT services [8].

Research on wireless sensor networks has been done by Calvo-Fullana et al. A MILP-based sensor selection method that takes energy harvesting and signal quality into account was presented in 2016. In more than 40 % of sensors, the EH-aware method achieved better reconstruction distortion than EH-agnostic methods. However, the use of fictitious datasets and strict assumptions limits the practical usefulness [9].

Xu et al. concentrate on optimal placement and scenario-based selection. A technique for choosing energy-efficient sensors in medical footwear was created in 2015. Although it successfully reduces energy usage, its applicability is limited by the absence of real-world proof. To improve feasibility, more research is required [10].

Lin and associates presented a multi-criteria method for choosing Internet of Things sensors. In 2017, the emphasis was on improving service quality, protecting the environment, and increasing energy efficiency. Although simulations demonstrated effective optimisation strategies, oversimplification assumptions and inadequate validation limit the practical implementation. To comprehend the limitations in practical situations, more research is required [11].

MOEA/D-OBL's performance was greatly enhanced with the addition of OBL (2014) to multi-objective optimisation, leading to quicker convergence and better solution quality. More research is needed to address issues like parameter selection and processing costs in order to improve practicality and reliability [12].

Wang et al. first presented the Opposition Krill Herd (OKH) algorithm. In 2016, it combined opposition-based learning with Cauchy mutation. When compared to alternative methods, OKH showed better performance in terms of efficiency, quality, and convergence. It is advised to investigate scalability and practical implementation in further detail [13].

The study carried out by Feng and associates. OBMBO-GP, a method that successfully solved challenging 0-1 Knapsack Problems by combining opposition-based learning, monarch butterfly optimisation, and Gaussian perturbation, was the main emphasis in 2016. It outperforms seven algorithms, but more study is needed to expand its usefulness [14]. Aziz and associates. A PSO-Voronoi method was put forth in 2019 to improve WSN coverage optimisation. In the end, it led to increased coverage and better efficiency while assuming consistent sensors and locations. To investigate adaption strategies and confirm their usefulness, more research is required [15].

Handy and coworkers. In 2002, deterministic cluster head selection was implemented to improve LEACH for wireless sensor networks, resulting in a 30 % increase in network lifetime. Energy use at base stations was underestimated, highlighting the necessity for further investigation of energy-efficient gearbox methods [16].

Chen and colleagues. In 2020, a dynamic clustering model for VANETs was created based on connection predictions. It lowered latency and enhanced delivery, albeit only with simulated data. Future research should focus on scaling concerns and limitations in real-world contexts [17].

Mnasri and companions. In 2020, a hybrid IoT optimisation technique was created to improve coverage, connectivity, and endurance when establishing 3D networks. The constrained indoor compatibility emphasises the necessity for more validation and expansion study [18].

Zhang and colleagues. The year is 2020. Enhanced MOEA/D with an information feedback model (MOEA/D-IFM) has resulted in improved solution quality and convergence. More testing and research on parameter sensitivity are required to improve reliability. Please reword this statement more smoothly [19].

Gu and colleagues. In 2020, improvements were made to NSGA-III by adding IFM, resulting in increased variety and convergence. On the other hand, further research is needed due to the limited assessment of different methodologies and the requirement for scalability [20].

Li and colleagues did this investigation. In 2019, we looked at over 100 quality indicators to evaluate solution sets in multiobjective optimisation, focusing on factors such as capacity, convergence, diversity, and convergence-diversity. Factors such as GD and IGD were investigated, but the lack of an ideal standard made the assessment difficult. Future study intends to help with the selection of acceptable indicators and their practical use [21].

Rahnamayana and colleagues did this study. Opposition-based sampling (OBS) was launched for optimisation in 2012, demonstrating its superiority to random sampling in terms of delivering optimal solutions. Despite the benefits of scalability, it is crucial to emphasise that problems may occur due to the reliance on certain assumptions and the uncertainty associated with projecting solution distances. Additional research is required for practical application [22].

Li et al. Parallel coordinates were proposed in 2017 as an effective approach for visually depicting sets of multi-objective solutions, which can help with decision-making and interpreting complex multidimensional data. Difficulties such as data loss and misinterpretation require further exploration to improve visualisation accuracy and user comprehension. Important grading criteria are convergence, coverage, homogeneity, and scope [23].

7. Results

Although there have been significant advances, the approaches and algorithms presented for sensor selection in IoT systems still have limits, as do the optimization difficulties they imply. Evolutionary algorithms such as MOEA/D and NSGA-III have demonstrated exceptional performance. Nonetheless, practical implementation limits. processing complexities, and scalability concerns persist. However, they also raised concerns about parameter biases and their suitability for specific domains. Energy optimization approaches such as utility-based sensor selection and energy-efficient mapping have demonstrated considerable benefits in terms of energy usage and network lifetime. However, scalability and application of these solutions in real-world circumstances are challenging.

Network performance and coverage efficiency were improved by the use of optimization and coverage algorithms that employed techniques such as PSO and dynamic clustering. Nonetheless, these approaches have some limitations, such as their unsuitability for varied sensor settings and reliance on simulated data. According to research, there are persistent challenges in assessing solution sets and adjusting to various scenarios within the boundaries of quality evaluation in multi-objective optimization and multi-population techniques.

Despite the loss of information caused by lowering dimensionality and assuming uniformly distributed solutions, techniques such as opposition-based sampling and parallel coordinates in visualization and sampling provided useful insights into multi-objective solution sets. Overall, while these methods have achieved significant advances, more study is needed to overcome their limitations and improve their applicability in practical scenarios.

8. Conclusions

The review paper provides a complete overview of multi-objective evolutionary algorithms used to select sensor sites in IoT systems, drawing on research findings from a variety of domains and actual applications. The findings emphasized the relevance of improving algorithms to meet complicated optimization difficulties, as well as the need for ongoing research to remove existing barriers and explore into unknown territory within the field.

References

[1]. I. Younas, A. Naeem, Optimization of sensor selection problem in IoT systems using opposition-based 7th International Conference on Advances in Signal Processing and Artificial Intelligence (ASPAI' 2025), 8-10 April 2025, Innsbruck, Austria

learning in many-objective evolutionary algorithms, *Comput. Electr. Eng.*, Vol. 97, 2022, 107625.

- [2]. F. Bian, D. Kempe, R. Govindan, Utility-based sensor selection, in *Proceedings of the Fifth Int. Conference Inf. Process. Sens. Networks (IPSN'06)*, 2006, pp. 11-18.
- [3]. L. P. Damuut, D. Gu, A mixed genetic algorithm strategy to sensor selection problem in WSNs, in *Proceedings of the 5th Int. Conference Comput. Intell. Commun. Syst. Networks (CICSyN'13)*, 2013, pp. 94-100.
- [4]. Q. Zhang, H. Li, MOEA/D: A multiobjective evolutionary algorithm based on decomposition, *IEEE Trans. EVol. Comput.*, Vol. 11, Issue 6, 2007, pp. 712-731.
- [5]. K. Deb, H. Jain, An evolutionary many-objective optimization algorithm using reference- point-based nondominated sorting approach, Part I: Solving problems with box constraints, *IEEE Trans. Evol. Comput.*, Vol. 18, Issue 4, 2014, pp. 577-601.
- [6]. H. Ma, S. Shen, M. Yu, Z. Yang, M. Fei, H. Zhou, Multi-population techniques in nature inspired optimization algorithms: A comprehensive survey, *Swarm EVol. Comput.*, Vol. 44, 2018, pp. 365-387.
- [7]. Z. Huang, K. J. Lin, L. Han, An energy sentient methodology for sensor mapping and selection in IoT systems, in *Proceedings of the IEEE Int. Symposium Ind. Electron.*, 2014, pp. 1436-1441.
- [8]. S. Y. Yu, C. S. Shih, J. Y. J. Hsu, Z. Huang, K. J. Lin, QoS oriented sensor selection in IoT system, in Proceedings of the IEEE Int. Conference Internet Things (iThings'14), IEEE Int. Conference Green Comput. Commun (GreenCom'14), IEEE Int. Conference Cyber-Physical-Social Comput (CPS'20), 2014, pp. 201-206.
- [9]. M. Calvo-Fullana, J. Matamoros, C. Anton-Haro, Sensor selection in energy harvesting wireless sensor networks, in *Proceedings of the IEEE Glob. Conference Signal Inf. Process (Glob'15)*, Vol. 1, 2016, pp. 43-47.
- [10]. T. Xu, M. Potkonjak, Energy saving using scenario based sensor selection on medical shoes, in *Proceedings of the IEEE Int. Conference Healthc. Informatics (ICHI'15)*, 2015, pp. 398-403.
- [11]. C.-C. Lin, D.-J. Deng, L.-Y. Lu, National Chiao Tung University Many-objective sensor selection in IOT system, *IEEE Wirel. Commun.*, Vol. 24, Issue 3, June 2017, pp. 40-47.

- [12]. X. Ma et al., MOEA/D with opposition-based learning for multiobjective optimization problem, *Neurocomputing*, Vol. 146, 2014, pp. 48-64.
- [13]. G. G. Wang, S. Deb, A. H. Gandomi, A. H. Alavi, Opposition-based krill herd algorithm with Cauchy mutation and position clamping, *Eurocomputing*, Vol. 177, 2016, pp. 147-157.
- [14]. Y. Feng, G. G. Wang, J. Dong, L. Wang, Oppositionbased learning monarch butterfly optimization with Gaussian perturbation for large-scale 0-1 knapsack problem, *Comput. Electr. Eng.*, Vol. 67, 2018, pp. 454-468.
- [15]. N. A. B. Ab Aziz, A. W. Mohemmed, M. Y. Alias, A wireless sensor network coverage optimization algorithm based on particle swarm optimization and Voronoi diagram, in *Proceedings of the IEEE Int. Conference Networking, Sens. Control (ICNSC'09)*, 2009, pp. 602-607.
- [16]. M. J. Handy, M. Haase, D. Timmermann, Low energy adaptive clustering hierarchy with deterministic cluster-head selection, in *Proceedings of the 4th Int. Workshop Mob. Wirel. Commun. Network* (MWCN'02), 2002, pp. 368-372.
- [17]. J. Cheng, G. Yuan, M. Zhou, S. Gao, Z. Huang, C. Liu, A connectivity-prediction-based dynamic clustering model for VANET in an urban scene, *IEEE Internet Things J.*, Vol. 7, Issue 9, 2020, pp. 8410-8418.
- [18]. S. Mnasri, N. Nasri, M. Alrashidi, A. van den Bossche, T. Val, IoT networks 3D deployment using hybrid many-objective optimization algorithms, *J. Heuristics*, Vol. 26, Issue 5, 2020, pp. 663-709.
- [19]. Y. Zhang, G. G. Wang, K. Li, W. C. Yeh, M. Jian, J. Dong, Enhancing MOEA/D with information feedback models for large-scale many-objective optimization, *Inf. Sci. (NY).*, Vol. 522, 2020, pp. 1-16.
- [20]. Z. M. Gu, G. G. Wang, Improving NSGA-III algorithms with information feedback models for large-scale many-objective optimization, *Futur. Gener. Comput. Syst.*, Vol. 107, 2020, pp. 49-69.
- [21]. M. Li, X. Yao, Quality evaluation of solution sets in multiobjective optimisation: A survey, ACM Comput. Surv., Vol. 52, Issue 2, 2019, 26.
- [22]. S. Rahnamayan, G. G. Wang, M. Ventresca, An intuitive distance-based explanation of oppositionbased sampling, *Appl. Soft Comput. J.*, Vol. 12, Issue 9, 2012, pp. 2828-2839.
- [23]. M. Li, L. Zhen, X. Yao, How to read many-objective solution sets in parallel coordinates [educational forum], *IEEE Comput. Intell. Mag.*, Vol. 12, Issue 4, 2017, pp. 88-100.

(011)

Soft Computing for Flood Susceptibility Mapping of Kullu District of India

<u>Shweta Vincent</u>¹, Mahesh Anil Inamdar¹, Om Prakash Kumar², Rohit Narayan H S¹, Nakul Rajendra Varma¹, Kaushik Naidu¹ and Anadya Dang²

¹ Department of Mechatronics, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, 576104, Karnataka, India

² Department of Electronics and Communication Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, 576104, Karnataka, India E-mail: omprakash.kumar@manipal.edu

Summary: Flood susceptibility mapping is vital in mitigating flood risks and managing disaster preparedness. This article presents the usage of machine learning models such as Random Forest (RF), Support Vector Machine (SVM), Gradient Boosting, Artificial Neural Networks (ANN), K- Nearest Neighbours (KNN) and Decision Tree (DT) for classification of flood-prone areas of the Kullu district in India. Ten factors were used to build the flood susceptibility model: slope, elevation, land use-land cover, normalized differences vegetation index (NDVI), topographical wetness index (TWI), drainage density (DD), distance to road and river, soil type, average rainfall, and flood inventory data. The use of these machine learning models in mapping and assessing flood risk is examined, and recommendations for further study are made. The model's output has been assessed using the AUC-ROC method and it has been observed that Gradient Boosting performs the best out of all models with an AUC score of 0.96.

Keywords: Flood susceptibility mapping (FSM), Flood conditioning factors (FCF), Flood inventory mapping (FIM), Machine learning, Confusion matrix.

1. Introduction

Flood susceptibility mapping is paramount for effective disaster risk management, particularly in regions like Manali, India, prone to flooding due to complex terrain and hydrological dynamics. While complete flood control remains elusive, flood susceptibility maps (FSMs) offer invaluable insights for managing pre- and post-flood scenarios amidst climatic and human-induced challenges. Coupled with the ever-growing global population, increasingly frequent and intense rainfall events due to climate change further exacerbate flood risks, potentially impacting millions.

To address this critical need, this study explores the development of FSMs using a multi-criteria decision support system tailored for the Manali region of North East India. Leveraging machine learning, the study aims to enhance flood prediction accuracy and delineate areas susceptible to flooding. Information gain theory will identify critical flood conditioning factors to obtain a robust flood risk map.

The research methodology involves analyzing geo-environmental parameters like elevation, slope, land use/land cover, and hydrological indices, integrating historical flood data to train and validate the FSM model. The effectiveness of the algorithms in generating FSMs will be rigorously evaluated using confusion matrix parameters and the AUC-ROC method.

1.1. The Problem of Floods and the Objective

Traditional flood susceptibility mapping in Kullu District relies on limited data and fails to capture the

complex interplay of factors like urbanization, coastal location, and changing rainfall patterns. These limitations lead to inaccurate flood maps, hindering effective flood preparedness and risk management strategies. Kullu's recent floods highlight the urgent need for improved flood prediction. Existing maps lack the necessary precision to guide targeted mitigation efforts. Traditional methods struggle to integrate diverse geospatial data (soil properties, land cover, drainage) and account for the dynamic nature of floods influenced by rainfall intensity and duration. Additionally, they cannot adapt to changing environmental conditions caused by climate change.

The objective of our research is to leverage existing machine learning techniques to build a Flood Susceptibility Map (FSM) for Kullu. The effectiveness of the generate map would be determined by comparing it with various other maps generated by other machine learning algorithms. Various Flood Conditioning Factors (FCFs) have been considered for the creation of the final FSMs.

1.2. Organization of This Article

The next section of this article presents a brief literature review of the various techniques available for the creation of FSMs and their corresponding accuracies. The third section of our article outlines our general methodology and further the fourth section presents the results and discussion. Finally, the fifth section of the article concludes the article and presents the future scope for study.

2. Literature Review

Floods are a major natural hazard causing significant loss of life and economic damage. Flood prediction methods are crucial for mitigating these impacts. This review examines the literature on flood risk mapping using Geographic Information Systems (GIS) and multi- criteria analysis. GIS plays a vital role in flood susceptibility mapping by enabling the analysis of spatial creation and data on flood-conditioning factors. Multi-criteria analysis techniques are employed to integrate these factors and assess flood susceptibility. The selection of appropriate flood-conditioning factors is crucial. Common factors include topography (slope, elevation), hydrology (drainage density, stream network), land cover, and soil characteristics. Studies [3, 10, 18] emphasize selecting factors considering the study area and availability of data.

Various methods are used for flood susceptibility mapping. These models assess the relationship between individual flood- conditioning factors and flood occurrences. Examples include Frequency Ratio (FR) and Information Value (IV) models used by [3, 14]. These techniques, like the Analytical Hierarchy Process (AHP), allow for incorporating expert knowledge and assigning weights to different

factors. Studies by [2, 9, 16] demonstrate this approach. Techniques like Support Vector Machines (SVM), Random Forest, and Artificial Neural Networks are increasingly being used for flood susceptibility mapping due to their ability to handle complex relationships between factors [6, 7, 19]. Integration of remote sensing data, particularly from Synthetic Aperture Radar (SAR), can enhance flood susceptibility mapping by providing information on water presence and inundation extent [1]. Studies are exploring novel approaches like a combination of decision table classifiers with metaheuristic algorithms [11] and Explainable Artificial Intelligence (XAI) models for improved transparency and interpretability [4].

[3, 9] emphasize the use of AHP for FSM creation. [14, 17] demonstrate the application of FR and other statistical models. Additional References: [4, 6, 7], [11, 15, 19] explore ML and advanced techniques for flood risk mapping. [5, 12] discuss the use of geospatial analysis and multivariate statistical methods. Table 1 outlines a summary of the various soft computing techniques used in state-of-the-art FSM creation. The table also showcases the parameters of evaluation used for the creation of the FSMs and their validation.

Reference No.	Algorithm/ Technique used	Evaluation Metrics		Performa	ice score	
	Support Vector Machine (SVM)	Confusion Matrix	Metric	PSO	SVM	GA
[1]	Particle Swarm Optimization (PSO)	Cohen Kappa	Accuracy	0.875	0.853	0.891
[1]	Genetic Algorithm (GA)	F1-score	F1 score	0.874	0.852	0.890
	Ensemble techniques	AUC-ROC	Kappa	0.750	0.706	0.782
[2]	Weighted criteria by pair-wise	Consistency ratio		CR =	0.8	
[2]	comparison matrix in AHP	AUC-ROC		AUC =	0.711	
[3]	Frequency ratio	AUC-ROC		AUC =	0.916	
[5]	Bivariate Statistical Model	лос-кос		AUC	0.910	
[4]	[4] Morphometric Analysis AUC-ROC		AUC =	0.892		
	K Nearest Neighbour (KNN)	Confector Matrix	Metric	KNN	SVM	ANN
[5]	Support Vector Machine (SVM)		Accuracy	0.915	0.927	0.927
	Artificial Neural Network (ANN)	AUC-KUC	AUC	0.913	0.943	0.934

Table 1. Literature Review.

From the review it is clear that, there are several authors who have explored the usage of the machine learning algorithms for the creation of FSMs. Nevertheless, our article presents the FSM creation for a flood-prone district of India, which is Kullu.

3. Methodology

This section outlines the methodology used for the creation of the FSM of Kullu. As outlined in Fig. 1, the Flood Inventory Map (FIM) and the Flood Conditioning Factors (FCF) are created in the following steps.

Data is obtained from various sources of field study, Google Earth and USGS Earth Explorer to create the FIM and FCFs of Kullu district. The Landsat 8 data with a spatial resolution of 30m has been used for this study. The various FCFs which have been generated using the ArcGIS tool and the Digital Elevation Model (DEM) of Kullu are, Slope, Elevation, Land Use Land Cover (LULC), Normalized Difference Vegetation Index (NDVI), Topographical Wetness Index (TWI), Drainage Density (DD), Distance to road and rive, Soil type and Average rainfall.

Each of these thematic maps are overlayed one over the other to obtain the data points. The FIM contains the information of the historical floods. This is overlayed over the FCF maps and hence a comprehensive dataset for training and testing is obtained.

Fig. 2, showcases the usage of these two maps i.e. the FIM and FCF for the final creation of the FSM. Figs. 3, 4, 5, and 6 describe the results obtained.

7th International Conference on Advances in Signal Processing and Artificial Intelligence (ASPAI' 2025), 8-10 April 2025, Innsbruck, Austria



Fig. 1. Creation of FIM and FCF.

The FIM and the FCF are given as inputs to the various machine learning algorithms and in turn, the FSM is obtained. The machine learning algorithms may/may not be fine-tuned for weights using metaheuristics (not used in our study). Once the model

has been created, it is validated and the final AUC-ROC scores are generated.



Fig. 2. Creation of FSM.



Fig. 3. FCFs generated for Kullu.

7th International Conference on Advances in Signal Processing and Artificial Intelligence (ASPAI' 2025), 8-10 April 2025, Innsbruck, Austria



Fig. 4. Final FSM generated for Kullu using Gradient Boosting with highest AUC score.



Fig. 5. AUC-ROC analysis.



Fig. 6. Comparative performance of ML algorithms.

4. Results and Discussion

Based on the aforementioned methodology, the various FCFs which have been created are showcased in Fig. 3. The final FSM generated for the Gradient Boosting algorithm has been showcased in Fig. 4. The comparative AUC-ROC curves for the various machine learning algorithms are shown in Fig. 5 and the comparative performance scores based on the metrics of the confusion matrix are shown in Fig. 6.

The AUC-ROC curve scores show that the Gradient Boosting algorithm outperforms all the other algorithms i.e. RF, SVM, ANN, KNN and DT. The final FSM created shows that the regions along the major water-ways i.e. rivers are most prone to flooding.

5. Conclusion and Future Scope

Floods are one of the most troubling natural disasters which lead to loss of life and property. It is extremely important that government authorities are able to predict floods earlier based on various factors in order to minimize the risk to life and property. One such technique is to create FSMs for a region based on pre-conditioning factors, i.e. FCFs. This article presented the usage of machine learning techniques for the creation of an FSM for the Kullu district of India.

It was observed, that the Gradient Boosting algorithm was able to generate the FSM with the highest AUC score of 0.96. Though all other algorithms performed almost at par with the Gradient Boosting algorithm, they are not considered in detail as the spatial resolution of the maps is 30 m, which means that a very high accuracy of the FSM would lead to maximum safety.

In the future, this research will be fine-tuned using metaheuristic algorithms to tune the classification weights of the algorithms, and further refine the accuracy of the maps.

References

- A. Mohammadi, K. V. Kamran, S. Karimzadeh, H. Shahabi, N. Al Ansari, Flood detection and susceptibility mapping using Sentinel-1 time series, alternating decision trees, and Bag-ADTree models, *Complexity*, Vol. 2020, 2020, 4271376.
- [2]. M. O. Faizan, M. Hudait, K. Sengupta, B. Roy Chowdhury, Flood susceptibility mapping using MCDM-AHP approach and geospatial techniques – a study of Hyderabad District, Telangana, India, *International Journal of Hydrology Science and Technology*, Vol. 16, 2023, pp. 204-221.
- [3]. A. Addis, GIS-based flood susceptibility mapping using frequency ratio and information value models in upper Abay River basin, Ethiopia, *Natural Hazards Research*, Vol. 3, Issue 2, June 2023, pp. 247-256.
- [4]. B. Pradhan, S. Lee, A. Dikshit, H. Kim, Spatial flood susceptibility mapping using an explainable artificial

intelligence (XAI) model, *Geoscience Frontiers*, Vol. 14, Issue 6, November 2023, 101625.

- [5]. K. Plataridis, Z. Mallios, Flood susceptibility mapping using hybrid models optimized with Artificial Bee Colony, *Journal of Hydrology*, Vol. 624, September 2023, 129961.
- [6]. A. Salvati, et al., Flood susceptibility mapping using support vector regression and hyper-parameter optimization, *Journal of Flood Risk Management*, Vol. 16, Issue 4, 2023, e12920.
- [7]. M. Ahmadlou, et al., Flood susceptibility mapping and assessment using a novel deep learning model combining multilayer perceptron and autoencoder neural networks, *Journal of Flood Risk Management*, Vol. 14, Issue 1, March 2021, e12683.
- [8]. A. Habibi, M. R. Delavar, M. S. Sadeghian, B. Nazari, Flood susceptibility mapping and assessment using regularized random forest and naive bayes algorithms, *ISPRS Annals of Photogrammetry, Remote Sensing* and Spatial Information Sciences, Vol. X-4/W1, 2023.
- [9]. K. C. Swain, C. Singha, L. Nayak, Flood susceptibility mapping through the GIS-AHP technique using the cloud, *ISPRS International Journal of Geo-Information*, Vol. 9, Issue 12, 2020, 720.
- [10]. B. H. Narendra, et al., Flood susceptibility mapping based on watershed geomorphometric characteristics and land use/land cover on a small island, *Global Journal of Environmental Science and Management*, Vol. 10, Issue 1, 2024, pp. 301-320.
- [11]. S. Askar, et al., Flood susceptibility mapping using remote sensing and integration of decision table classifier and metaheuristic algorithms, *Water*, Vol. 14, Issue 19, 2022, 3062.
- [12]. A. Ahmed, et al., Flood susceptibility mapping utilizing the integration of geospatial and multivariate statistical analysis, Erbil area in Northern Iraq as a case study, *Scientific Reports*, Vol. 13, 2023, 11919.
- [13]. W. Fenglin, et al., Flood susceptibility mapping using machine learning algorithms: A case study in Huong Khe District, Ha Tinh Province, Vietnam, *Scientific Reports*, Vol. 13, 2023.
- [14]. S. I. Majid, M. Kumar, P. Kumar, N. K. Verma, GIS based flood susceptibility mapping of Srinagar District, India using Weights of Evidence (WofE), Frequency Ratio (FR) and Fuzzy Gamma Operator (FGO), Journal of the Indian Society of Remote Sensing, Vol. 51, 2023, pp. 2421-2446.
- [15]. A. Arabameri, et al., Flood susceptibility mapping using meta-heuristic algorithms, *Geomatics, Natural Hazards and Risk*, Vol. 13, 1, 2022, pp. 949-974.
- [16]. N. N. B. Khairul Anuar, Flood susceptibility mapping using GIS and AHP in Kelantan, MS Thesis, University of Kelantan, Kelantan, 2022.
- [17]. Z. U. Rahman, et al., GIS-based flood susceptibility mapping using the bivariate statistical model in Swat River Basin, Eastern Hindukush region, Pakistan, *Frontiers in Environmental Science*, Vol. 11, 2023.
- [18]. M. Hasanuzzaman, et al., Flood susceptibility mapping using morphometric parameters and GIS, in Spatial Modelling of Flood Risk and Flood Hazards, Pradhan, B., et al. (Eds.), *Springer*, 2022, pp. 15-31.
- [19]. D. L. Nguyen, et al., Flood susceptibility mapping using machine learning algorithms: a case study in Huong Khe District, Ha Tinh Province, Vietnam, *International Journal of Geoinformatics*, Vol. 19, Issue 7, 2023, pp. 1-15.

(014)

Rényi Entropy-based Shrinkage Algorithm for Sparse Time-frequency Distribution Reconstruction Using Component Alignment Map

V. Jurdana

University of Rijeka, Faculty of Engineering, Department of Automation and Electronics, Vukovarska 58, 51000 Rijeka, Croatia Tel.: + 385 51 505 660 E-mail: vedran.jurdana@riteh.uniri.hr

Summary: Time-frequency distributions (TFDs) are powerful tools for analyzing non-stationary signals, providing insightful representations of their time-varying spectral content. Compressive sensing (CS) has emerged as an advanced technique in this field, enabling the reconstruction of TFDs from sparse ambiguity function samples. Despite its high performance, a key challenge lies in selecting the optimal regularization parameter. A Rényi entropy-based shrinkage algorithm was proposed to address this, utilizing the local Rényi entropy (LRE) and estimated local component counts to achieve more interpretable and accurate TFD shrinkage compared to conventional thresholding approaches. However, the algorithm's performance is constrained by the inherent limitations of LRE. This paper presents a novel enhancement by integrating the component alignment map (CAM), which identifies and isolates regions of the TFD with similar components. CAM improves local component estimation, refines the shrinkage process, and reduces algorithmic parameters. Experimental results demonstrate the enhanced algorithm's superior reconstruction performance for synthetic and real-world electroencephalogram signals, outperforming existing shrinkage algorithm.

Keywords: Time-frequency distribution, Compressive sensing, Local Rényi entropy, Signal reconstruction, Multi-objective optimization.

1. Introduction

Time-frequency distributions (TFDs) are essential for analyzing non-stationary signals [1, 2]. While linear methods are inherently limited by the Heisenberg uncertainty principle, quadratic TFDs (QTFDs) often suffer from cross-term interference, which can obscure true signal components, commonly referred to as auto-terms [1]. To address these limitations, non-linear methods have been developed, such as synchroextracting transform (SET) [3] and CS-based methods, demonstrating robust performance across diverse signal types [4-7].

The CS technique explored in this study reconstructs TFDs from a subset of samples in the ambiguity function (AF). Importantly, the objective of CS-based methods in time-frequency signal analysis is to reconstruct the signal's auto-terms rather than the complete initial TFD [6, 7]. A key challenge in these methods is determining the optimal regularization parameter. If the regularization parameter is set too high, the reconstructed TFD can become overly sparse, leading to incomplete auto-terms [6, 7].

To overcome time-consuming process of manual or experimental selections of the regularization parameter, previous work [8] proposed TFD regularization by utilizing local Rényi entropy (LRE) [9, 10]. The usage of the estimated local numbers of components from the LRE enabled more effective TFD shrinkage on auto-terms compared to conventional regularization parameter [7, 8, 11, 12].

This paper improves the Rényi entropy-based shrinkage algorithm by integrating the component alignment map (CAM) introduced in [13]. CAM effectively distinguishes TFD regions based on the alignment of components, identifying whether they are more oriented toward the time axis or the frequency axis. Components aligned with the time axis are better analyzed using time slices, while those aligned with the frequency axis are more effectively processed with frequency slices [13]. By leveraging CAM's ability to group components with similar alignments, the Rényi entropy-based shrinkage algorithm is simplified and utilizes more precise LRE estimates.

The superior reconstruction accuracy of this improved algorithm is demonstrated on both synthetic and real-world electroencephalogram (EEG) seizure signals, which are characterized by multiple components with distinct alignments along the time and frequency axes.

2. Time-frequency Signal Analysis

The Wigner-Ville distribution (WVD) is a widely used TFD defined as [1]:

$$W_{Z}(t,f) = \int_{-\infty}^{\infty} z\left(t + \frac{\tau}{2}\right) z^{*}\left(t - \frac{\tau}{2}\right) e^{-j2\pi f\tau} d\tau, \quad (1)$$

where z(t) denotes the analytic form of a multi-component, non-stationary signal. The WVD is highly effective for estimating the instantaneous frequency of single-component linear frequency modulation (LFM) signals. However, for multi-component signals, the WVD introduces cross-terms, which obscure the representation and limit its applicability in modern signal analysis [1]. To mitigate cross-terms, the AF is employed, defined as [1]:

$$A_z(\nu,\tau) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} W_z(t,f) e^{j2\pi(f\tau-\nu t)} dt df \qquad (2)$$

Cross-terms, being highly oscillatory, are positioned away from the AF origin, unlike auto-terms, which are concentrated along the AF origin trajectory [1]. Consequently, 2D low-pass filtering in the AF domain is commonly used to suppress cross-terms. This leads to the class of QTFDs, $\rho_z(t, f)$, defined as [1]:

$$\rho_z(t,f) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} A_z(v,\tau) g(v,\tau) e^{j2\pi(vt-f\tau)} dv d\tau, \qquad (3)$$

where $g(v, \tau)$ is the low-pass filter applied in the AF. Although QTFDs improve cross-term suppression, they inherently involve a trade-off between auto-term resolution and cross-term reduction [1, 2].

To address these limitations, the adaptive directional TFD (ADTFD) incorporates adaptive filtering by adjusting the smoothing direction at each time-frequency (TF) point as [14]:

$$\rho_{(ad)}(t,f) = \rho_z(t,f) ** \gamma_\theta(t,f) \tag{4}$$

Here, the double asterisk denotes double convolution in t and f, and $\gamma_{\theta}(t, f)$ is the smoothing kernel [14]:

$$\gamma_{\theta}(t,f) = \frac{ab}{2\pi} \frac{d^2}{df_{\theta}^2} e^{-a^2 t_{\theta}^2 - b^2 f_{\theta}^2}, \qquad (5)$$

whose direction is controlled by θ , while $t_{\theta} = t \cos(\theta) + f \sin(\theta)$ and $f_{\theta} = -t \sin(\theta) + f \cos(\theta)$. Parameters *a* and *b* control the smoothing along the time and frequency axes, respectively [14].

To automatically optimize the ADTFD, the locally adaptive ADTFD (LO-ADTFD) is employed [15, 16]. This approach selects TF points by minimizing across a predefined set of ADTFDs. The parameter set $(a, b) = \{(2,20), (2,30), (3,6), (3,8)\}$ is used, while the smoothing window length of $\gamma_{\theta}(t, f)$ is optimized for each combination as detailed in [15, 16].

2.1. Compressive Sensing-based Method

Further improvements in TFD performance have been achieved by leveraging CS method [4-7]. The CS-based method utilized in this study reconstructs TFD from a sparse subset of AF samples, referred to as the CS-AF region $A_z^{CS}(v,\tau)$. Given that the CS-AF area resembles a rectangle centred at the AF origin similar to low-pass filtering, reconstruction algorithms are employed to iteratively improve the loss in auto-term resolution [4-7]. In this study, an adaptive CS-AF area with size $N'_{\tau} \times N'_{\nu}$ is used, where N'_{τ} and N'_{ν} are the numbers of lag and Doppler bins, respectively [17]. Importantly, the CS-AF area must only include auto-term samples; otherwise, interference may reappear in the reconstructed TFD.

The reconstruction algorithm seeks the optimal TFD solution by solving [4, 6, 7]:

$$\mathbf{Y}_{\mathbf{z}}(t,f) = \mathbf{\Psi}^{H} \cdot \mathbf{A}_{\mathbf{z}}^{CS}(\nu,\tau), \tag{6}$$

where Ψ^{H} is the Hermitian transpose of a domain transformation matrix. Since multiple solutions for $\Upsilon_{z}(t, f)$ are possible, the regularization function emphasizes the desired properties of the solution [4, 6, 7]. The l₁ norm is usually employed to promote sparsity, leading to the unconstrained optimization problem [4, 6, 7, 18, 19]:

$$\begin{aligned} \mathbf{Y}_{\mathbf{z}}^{\mathbf{l}_{1}}(t,f) &= \arg\min_{\mathbf{Y}_{\mathbf{z}}(t,f)} \left| \left| \mathbf{Y}_{\mathbf{z}}(t,f) \right| \right|_{1}, \\ \text{subject to:} \left| \left| \mathbf{Y}_{\mathbf{z}}(t,f) - \mathbf{\Psi}^{H} \mathbf{A}_{\mathbf{z}}^{CS}(\nu,\tau) \right| \right|_{2}^{2} \leq \epsilon, \end{aligned}$$
(7)

where ϵ represents the energy tolerance criterion. The closed-form solution when using the l_1 norm is given as [4, 6, 7, 18, 19]:

$$\mathbf{Y}_{\mathbf{z}}^{\mathbf{l}_{1}}(t,f) = \operatorname{soft}_{\lambda}\{\mathbf{Y}_{\mathbf{z}}(t,f)\},\tag{8}$$

where $\operatorname{soft}_{\lambda}\{Y_z(t, f)\} = \operatorname{sgn}(Y_z(t, f)) \max(|Y_z(t, f)| - \lambda, 0)$, with λ being the regularization parameter. The choice of λ is signal-dependent and non-trivial: a low λ value reconstructs interference with smeared auto-terms, while a high λ value results in the loss of critical auto-term components [4, 6, 7, 18, 19].

2.2. Measuring Sparse Time-frequency Distributions

A reliable and accurate metric for evaluating reconstructed TFDs is crucial. Computationally efficient global measures, which assess the TFD as a whole and yield a single output value, are commonly used. One such measure is the concentration metric [20]:

$$M = \frac{1}{N_t N_f} \left[\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left| \frac{\rho_z(t,f)}{\int_{-\infty}^{\infty} \rho_z(t,f) dt df} \right|^{\frac{1}{2}} dt df \right]^2, \tag{9}$$

where N_t and N_f represent the number of time samples and frequency bins, respectively. Another widely used global measure is the Rényi entropy, defined as [21]:

$$R = \frac{1}{1-\alpha_R} \log_2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(\frac{\rho_z(t,f)}{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \rho_z(t,f) dt df} \right)^{\alpha_R} dt df, \quad (10)$$

where α_R is usually chosen as an odd integer [21].

However, recent studies [7, 8, 11] have shown that these global measures are not adequate for assessing reconstructed TFDs since they do not provide information about the local positions of auto-terms and may treat auto-terms and interference equally. To address this, the LRE was implemented which captures the local behavior of components for each time, t_0 , and frequency slice, f_0 , as [7, 8, 11]:

$$NC_t^{\rho_z(t,f)}(t_0) =$$

= $2^{R(\chi_{t_0}\{\rho_z(t,f)\}) - R(\chi_{t_0}\{\rho_{\text{ref}}(t,f)\})},$ (11)

$$NC_{f}^{\rho_{z}(t,f)}(f_{0}) =$$

$$= 2^{R(\chi_{f_{0}}\{\rho_{z}(t,f)\}) - R(\chi_{f_{0}}\{\rho_{\text{ref}}(t,f)\})},$$
(12)

where notations *t* and *f* denote localization through time and frequency slices, respectively, while $\rho_{ref}(t, f)$ represents the reference TFD. The operators χ_{t_0} and χ_{f_0} extracts TFD samples within intervals $[t_0 - \Theta_t/2, t_0 + \Theta_t/2]$ and $[f_0 - \Theta_f/2, f_0 + \Theta_f/2]$, respectively, controlled with the window lengths Θ_t and Θ_f [7, 8, 11, 12].

The LRE enabled a comparison of the local number of components before and after reconstruction using the mean squared error (MSE) [8, 11, 12]:

$$MSE = \frac{1}{N_t} \sum_{t=1}^{N_t} \left(\frac{NC_t^{\rho_z(t,f)}(t) - NC_t^{\gamma_z^{\mathbf{l}_1}(t,f)}(t)}{\max\left(NC_t^{\rho_z(t,f)}(t), NC_t^{\gamma_z^{\mathbf{l}_1}(t,f)}(t)\right)} \right)^2 + \frac{1}{N_f} \sum_{f=1}^{N_f} \left(\frac{NC_f^{\rho_z(t,f)}(f) - NC_f^{\gamma_z^{\mathbf{l}_1}(t,f)}(f)}{\max\left(NC_f^{\rho_z(t,f)}(f), NC_f^{\gamma_z^{\mathbf{l}_1}(t,f)}(f)\right)} \right)^2$$
(13)

A high MSE indicates an oversparse TFD or one contaminated with interference, resulting in degraded auto-term quality.

2.3. Component Alignment Map

This paper leverages the CAM, a method proposed in [13], to segment the TFD into regions suitable for either time or frequency localization. The CAM is constructed through the following key steps [13]:

- 1. For the input TFD, two new TFDs are generated, each containing only auto-term maxima extracted based on the numbers of significant components NC_t (in time slices) and NC_f (in frequency slices);
- The connectivity of auto-terms in both TFDs is assessed using a metric that counts the number of connected regions of samples;
- 3. Based on the connectivity metric, auto-terms are classified as time-aligned (CAM(t, f) = 1) or frequency-aligned (CAM(t, f) = 0);
- 4. TFD segments with significant local maxima in NC_t or NC_f are extracted. Each segment undergoes further analysis through re-estimation of the local number of components using the LRE and re-evaluation using the connectivity metric from step 2;
- 5. The initial CAM from step 3 is refined for each segment identified in step 4 that contains components with differing alignments.

This creates CAM where ones indicate TFD regions where localization in time slices is needed, while zeros indicate regions where localization in frequency slices is preferred. By multiplying the CAM and TFD, components with similar alignments can be extracted. Two operators are then defined: $\eta_t \{\rho_z(t, f)\}$, which extracts components where CAM(t, f) = 1, and $\eta_f \{\rho_z(t, f)\}$, which extracts components where camponents where CAM(t, f) = 0 [13].

3. Rényi Entropy-based Shrinkage Algorithm

The Rényi entropy-based shrinkage algorithm, referred to as RTwIST [8, 11], builds on the two step iterative shrinkage/thresholding (TwIST) algorithm [22]. The update equation for the (n + 1)-th iteration is defined as [8, 11]:

$$\begin{bmatrix} \mathbf{Y}_{z}^{\mathbf{l}_{1}}(t,f) \end{bmatrix}^{[n+1]} = (1 - \alpha_{\mathrm{TwIST}}) \begin{bmatrix} \mathbf{Y}_{z}^{\mathbf{l}_{1}}(t,f) \end{bmatrix}^{[n-1]} + \\ + (\alpha_{\mathrm{TwIST}} - \beta_{\mathrm{TwIST}}) \begin{bmatrix} \mathbf{Y}_{z}^{\mathbf{l}_{1}}(t,f) \end{bmatrix}^{[n]} + \\ + \beta_{\mathrm{TwIST}} \mathrm{shrink}_{t,f} \begin{cases} \begin{bmatrix} \mathbf{Y}_{z}^{\mathbf{l}_{1}}(t,f) \end{bmatrix}^{[n]} + \\ + \mathbf{\Psi}^{H} \left(A_{z}^{CM}(\nu,\tau) - \mathbf{\Psi} \begin{bmatrix} \mathbf{Y}_{z}^{\mathbf{l}_{1}}(t,f) \end{bmatrix}^{[n]} \right) \end{cases}$$
(14)

Here, the shrink_{*t,f*} operator performs TFD shrinkage by retaining only the largest NC_t or NC_f surface areas around local maxima in time or frequency slices, respectively. These local maxima correspond to auto-terms. The amount of retained samples is controlled by parameters δ_t and δ_f , which adjust the sparsity. Lower $\delta_{t,f}$ values reduce oversparsity but also lower the resolution of the reconstructed TFD, while higher $\delta_{t,f}$ values improve resolution but may increase oversparsity [8, 11].

The result of shrink_{*t*,*f*} operator, denoted as $\varsigma_z(t, f)$, is obtained via iterative shrinkage over time and frequency slices, expressed as [8, 11]:

$$[\varsigma_z^{t,f}(t,f)]^{[n+1]} = \text{shrink}_{t,f}\{\varsigma_z'(t,f)\}, \quad (15)$$

where $\varsigma'_{z}(t,f) = [\mathbf{Y}_{z}^{l_{1}}(t,f)]^{[n]} + \Psi^{H} (A_{z}^{CM}(v,\tau) - \Psi[\mathbf{Y}_{z}^{l_{1}}(t,f)]^{[n]})$. In the existing algorithm, the outputs of the shrinkage operations, $\varsigma_{z}^{t}(t,f)$ and $\varsigma_{z}^{f}(t,f)$, are combined via a weighted average [8,11]:

$$[\varsigma_z(t,f)]^{[n+1]} = p \,\varsigma_z^{\ t}(t,f) + (1-p) \,\varsigma_z^{\ f}(t,f), \quad (16)$$

where p is the global weighting parameter in the range [0,1]. However, as p is applied globally across the entire TFD, it fails to account for components with locally varying directions and shapes. This limitation makes the use of a single p value inappropriate for signals with diverse local structures.

To address this issue, this paper introduces RTwIST-CAM, which replaces (16) by integrating the CAM as:

$$[\varsigma_z(t,f)]^{[n+1]} = \operatorname{shrink}_t\{\eta_t\{\varsigma'_z(t,f)\}\} + \\ +\operatorname{shrink}_f\{\eta_t\{\varsigma'_z(t,f)\}\}$$
(17)

The final shrunken TFD consists of two independent shrinkage operations, shrinkt and shrink_f, automatically guided by the CAM. Notably, the parameter p is no longer required, significantly simplifying the algorithm by reducing the number of parameters that end-users must specify for an unknown signal. Additionally, the multi-objective particle swarm optimization (MOPSO) method, as employed in [8], will optimize fewer parameters. This reduction simplifies the optimization problem and decrease the time required to achieve the optimal reconstructed TFD. Unlike the conventional RTwIST algorithm, which relies on predefined inputs NC_t and NC_f , RTwIST-CAM employs more precise local component estimates derived from TFDs with disjoint components: $\eta_t \{ \rho_z(t, f) \}$ and $\eta_f \{ \rho_z(t, f) \}$.

4. Experimental Results and Discussion

The performance of the proposed RTwIST-CAM, the existing RTwIST and SET algorithms was evaluated using both a synthetic signal, composed of six linear and non-linear components, and a real-world EEG seizure signal [7, 13, 14]. To enhance the EEG signal, a differentiator filter was applied to whiten the background and highlight spikes, as recommended in [23]. For the synthetic and EEG seizure signals, the CS-AF areas were calculated as 17 × 25 and 15 × 15, respectively, and the reconstruction parameter was fixed at $\epsilon = 10^{-3}$ as in [6-8,11-13]. The LRE was computed using LOADTFD and parameters $\alpha_R = 3$ with $\Theta_t = \Theta_f = 11$ for the synthetic signal, and $\Theta_t = \Theta_f = 5$ for the EEG seizure signal, following [13].

Fig. 1 illustrates the WVDs and LOADTFDs for the considered signals. It is evident that LOADTFD provides better representation for these signals, as cross-terms in the WVD obscure the true components.



Fig. 1. (a) WVD of synthetic signal; (b) WVD of EEG seizure signal; (c) LOADTFD of synthetic signal; (d) LOADTFD of EEG seizure signal.

Fig. 2 depicts the obtained CAMs for the considered signals. For the synthetic signal example, the black areas in CAM encompass two constant FM components, indicating their suitability for localization in time slices. Similarly, for the EEG seizure signal, the CAM distinguishes between the constant FM component and the spiky components.



Fig. 2. CAM for the considered signals: (a) synthetic signal;(b) EEG seizure signal. Areas in black represent zeros, while areas in white represent ones.

When estimating the local number of components from the entire TFD, artificial increase in the estimated component numbers can occur, as shown by the dotted lines in Fig. 3. These inaccuracies arise when the analyzed signal components deviate from the reference signal. This issue is mitigated by separating components with similar alignment through the multiplication of the CAM and the TFD, as illustrated in Fig 3. These refined estimates are subsequently incorporated into the RTwIST-CAM algorithm.



Fig. 3. Local number of components for the considered signals obtained using the LRE: (a,c) synthetic signal; (b,d) EEG seizure signal. Solid purple lines indicate estimations on the full TFD, while dotted red lines indicate estimations on separated TFDs using CAM.

Fig. 4 presents the obtained TFDs for the considered signals, with Table 1 providing quantitative performance metrics to complement the visual assessment. Note that the optimal RTwIST,

RTwIST-CAM, and SET parameters have been selected manually in this research.



Fig. 4. Obtained TFDs of the considered signals: (a) synthetic signal, SET; (b) EEG signal, SET; (c) synthetic signal, RTwIST ($\alpha_{TwIST} = 0.91$, $\beta_{TwIST} = 0.82$, $\delta_t = \delta_f = 0.85$, p = 0.7); (d) EEG signal, RTwIST ($\alpha_{TwIST} = 0.9$, $\beta_{TwIST} = 0.81$, $\delta_t = \delta_f = 0.92$, p = 0.6); (e) synthetic signal, RTwIST-CAM ($\alpha_{TwIST} = 0.91$, $\beta_{TwIST} = 0.8$, $\delta_t = \delta_f = 1$); (f) EEG signal, RTwIST-CAM ($\alpha_{TwIST} = 0.82$, $\delta_t = \delta_f = 1$).

Table1.PerformancecomparisonbetweentheRTwIST-CAM,RTwISTandSETalgorithmsfortheconsideredsignals.Valuesinboldindicatethebest-performing algorithm for each measure.

	Synthetic signal EEG seizure		Synthetic signal		signal	
	RTwIST	SET	RTwIST- CAM	RTwIST	SET	RTwIST- CAM
MSE	0.1478	0.1881	0.0878	0.1654	0.5587	0.1077
М	0.0177	0.0268	0.0103	0.0168	0.0895	0.0174

The reconstructed TFDs generated by the proposed RTwIST-CAM algorithm demonstrate superior auto-term resolution, continuity, and cross-term suppression compared to the RTwIST algorithm and SET. This improvement is particularly notable for the EEG seizure signal. That is, constant FM and spiky components are reconstructed with greater continuity and less non-linear distortions compared to the RTwIST algorithm, while the SET shows unsuitable for representing spikes.

The numerical results in Table 1 validate these observations. For the synthetic signal, both measures indicate significantly better performance using the proposed RTwIST-CAM algorithm. However, for the EEG seizure signal, the existing RTwIST algorithm shows better performance based on *M* measure. This results from discontinuities, such as missing auto-terms (as visible in Fig. 4d for RTwIST), which artificially reduce the M value. Nevertheless, the MSE measure highlights improved preservation of signal auto-terms when using the RTwIST-CAM algorithm.

The runtime complexity analysis of the RTwIST-CAM algorithm, based on 1000 independent simulation runs, reveals its performance characteristics. For synthetic and EEG signals, the LRE calculation times were 0.878 and 1.047 seconds, respectively, while CAMs were generated in 1.478 and 1.621 seconds. This indicates that RTwIST-CAM's input preparation time is approximately double that of RTwIST the original algorithm. However, RTwIST-CAM demonstrates faster convergence, reconstructing TFDs in 1.47 seconds compared to RTwIST's 1.66 seconds. The SET method processed signals more quickly (0.478s for synthetic, 0.587s for EEG), highlighting the higher computational complexity of CS-based methods. Notably, when using MOPSO for parameter optimization. CAM is calculated only once at the beginning, and the reduction in parameters requiring optimization is expected to significantly decrease the overall optimization time. The CS-based method is primarily designed for offline signal analysis applications.

4. Conclusions

This paper introduces the integration of the CAM with the RTwIST algorithm, specifically targeting multi-component signals with distinct alignments along the time and frequency axes of a TFD. Previous studies emphasize that components with different alignments require specific localization approaches. Failure to separate such components leads to less accurate LRE estimates, and, consequently, reduced performance of the original RTwIST algorithm.

The incorporation of CAM into the RTwIST algorithm offers several significant advantages. Firstly, it enables more accurate LRE estimates by separating components with distinct alignments Secondly, it facilitates more precise localization for the shrinkage operation by creating two TFDs with disjoint components of similar alignment. Thirdly, it automatically identifies which components should be analyzed with the appropriate localization and LRE estimates, effectively eliminating the need for the parameter p. This not only simplifies the manual selection of algorithm's parameters but also creates opportunities to reduce computational complexity when using meta-heuristic optimization methods for automatic parameter tuning.

Experimental validation using synthetic and real-world EEG seizure signals demonstrate the enhanced algorithm's superiority over the original RTwIST algorithm and SET, achieving better autoterm resolution, consistency, and cross-term suppression. This advancement strengthens CS methods over SET. Future research will investigate the effects of reducing the number of parameters when employing MOPSO and assess RTwIST-CAM's performance on noisy signals. Moreover, future work will explore algorithm's potential for real-time implementation, and the possibility of integrating deep learning and image processing techniques for CAM determination.

Acknowledgements

This research was supported by the University of Rijeka under the project number uniri-mladi-tehnic-23-2.

References

- B. Boashash, Time-Frequency Signal Analysis and Processing, A Comprehensive Reference, 2nd Ed., *Elsevier*, London, UK, 2016.
- [2]. L. Stankovic, M. Dakovic, T. Thayaparan, Time-Frequency Signal Analysis with Applications, *Artech House Publishers*, Boston, USA, 2013.
- [3]. G. Yu, M. Yu, C. Xu, Synchroextracting transform, *IEEE Transactions on Industrial Electronics*, Vol. 64, 2017, pp. 8042-8054.
- [4]. P. Flandrin, P. Borgnat, Time-frequency energy distributions meet compressed sensing, *IEEE Transactions on Signal Processing*, Vol. 58, Issue 6, 2010, pp. 2974-2982.
- [5]. L. Stanković, I. Orović, S. Stanković, M. Amin, Compressive sensing based separation of nonstationary and stationary signals overlapping in time-frequency, *IEEE Transactions on Signal Processing*, Vol. 61, Issue 18, 2013, pp. 4562-4572.
- [6]. I. Volarić, Signal concentration enhancement in the time-frequency domain using adaptive compressive sensing, PhD Thesis, Faculty of Engineering, *University of Rijeka*, 2017.
- [7]. V. Jurdana, A multi-objective optimization procedure for locally adaptive time-frequency analysis with application in EEG signal processing, PhD Thesis, Faculty of Engineering, *University of Rijeka*, 2023.
- [8]. V. Jurdana, I. Volaric, V. Sucic, Sparse time-frequency distribution reconstruction based on the 2D Renyi entropy shrinkage algorithm, *Digital Signal Processing*, Vol. 118, 2021, 103225.
- [9]. V. Sucic, N. Saulig, B. Boashash, Estimating the number of components of a multicomponent nonstationary signal using the short-term timefrequency Renyi entropy, *EURASIP Journal on*

Advances in Signal Processing, Vol. 2011, Issue 1, 2011, 125.

- [10]. V. Sucic, N. Saulig, B. Boashash, Analysis of local time-frequency entropy features for nonstationary signal components time supports detection, *Digital Signal Processing*, Vol. 34, 2014, pp. 56-66.
- [11]. V. Jurdana, I. Volaric, V. Sucic, The local Rényi entropy based shrinkage algorithm for sparse TFD reconstruction, in *Proceedings of the International Conference on Broadband Communications for Next Generation Networks and Multimedia Applications* (CoBCom'20), Graz, Austria, 2020, pp. 4-9.
- [12]. V. Jurdana, N. Lopac, M. Vrankic, Sparse time-frequency distribution reconstruction using the adaptive compressed sensed area optimized with the multi-objective approach, *MDPI Sensors*, Vol. 23, 2023, 4148.
- [13]. V. Jurdana, M. Vrankic, N. Lopac, G. M. Jadav, Method for automatic estimation of instantaneous frequency and group delay in time-frequency distributions with application in EEG seizure signals analysis, *MDPI Sensors*, Vol. 23, 2023, 4680.
- [14]. N. A. Khan, S. Ali, Classification of EEG signals using adaptive time-frequency distributions, *Metrol. Meas. Syst.*, Vol. 23, 2016, pp. 251-260.
- [15]. N. A. Khan, B. Boashash, Multi-component instantaneous frequency estimation using locally adaptive directional time frequency distributions, *Int. J. Adapt. Control. Signal Process.*, Vol. 30, 2016, pp. 429-442.
- [16]. M. Mohammadi, A. A. Pouyan, N. A. Khan, V. Abolghasemi, Locally optimized adaptive directional time-frequency distributions, *Circuits Syst. Signal Process.*, Vol. 37, 2018, pp. 3154-3174.
- [17]. I. Volaric, V. Sucic, S. Stankovic, A data driven compressive sensing approach for time-frequency signal enhancement, *Signal Processing*, Vol. 141, 2017, pp. 229-239.
- [18]. E. Sejdić, I. Orović, S. Stanković, Compressive sensing meets time frequency: An overview of recent advances in time-frequency processing of sparse signals, *Digital Signal Processing*, Vol. 77, 2018, pp. 22-35.
- [19]. Z. Zhang, Y. Xu, J. Yang, X. Li, D. Zhang, Survey of sparse representation: algorithms and applications, *IEEE Access*, Vol. 3, 2015, pp. 490-530.
- [20]. L. Stanković, A Measure of some time-frequency distributions concentration, *Signal Process.*, Vol. 81, 2001, pp. 621-631.
- [21]. R. G. Baraniuk, P. Flandrin, A. J. E. M. Janssen, O. J. J. Michel, Measuring time-frequency information content using the Renyi entropies, *IEEE Transactions* on *Information Theory*, Vol. 47, Issue 4, 2001, pp. 1391-1409.
- [22]. J. M. Bioucas-Dias, M. A. T. Figueiredo, A new TwIST: two-step iterative shrinkage/thresholding algorithms for image restoration, *IEEE Transactions on Image Processing*, Vol. 16, Issue 12, 2007, pp. 2992-3004.
- [23]. K. Majumdar, Differential operator in seizure detection, *Comput. Biol. Med.*, Vol. 42, 2012, pp. 70-74.

(015)

Applied AI for DLT and CLT with Imperfect Bonding

R. Hussein

State University of New York (emeritus), Syracuse, NY 13210, USA Tel.: + 001 315 695 2340 E-mail: ezpscg@gmail.com

Summary: Machine learning, deep learning, natural language processing, expert systems (ES), robotics, machine vision, and speech recognition are some of the subsets of artificial intelligence (AI). The ES subset, which has proved itself to be a potent tool for innovative engineering design and solutions, is the aim of this paper. To the best of our knowledge, there is no relevant ES (AI) literature on CLT and DLT, despite its glowing widespread use globally. This paper fills that void. The paper is divided into two sections. The evident inaccuracy of the DLT in some of the literature is examined in the first section. For the first time, bonding, interlayer slip, and ply angle are taken into consideration when deriving and solving the rigorous governing equations using Fourier trigonometric series. Essential characteristics like the ply angle, interlayer slips, and bonding stiffness are taken into account in the formulation for the first time. Both CLT and DLT can use the model and solution, which has been verified and validated. A parametric analysis is performed to ascertain how these properties affect structural performance. The findings show that when serviceability is the quantity of interest, bonding stiffness shouldn't be underestimated. Now, experts can quantify the ubiquitous, perfectly rigid bonding. The findings affirmed that panels subjected to transverse loads are better suited for CLT than DLT. The essential design characteristics were included in a practical formula for bending stiffness that was coined, verified, and validated. The second section introduces a novel toolkit and methodology based on deterministic ES. That makes it possible for virtual brains and human experts to collaborate to arrive at a satisfactory design.

Keywords: Engineered timber laminates, Mass timber, Bending stiffness, Bonding stiffness, CLT, DLT, GLT, Interlayer slip.

1. Introduction – Hankinson's Formula

In general, wood composites date back to ancient Egypt [1]. The first patent application for the new cross-laminated timber, or CLT, was submitted in 1920 [2-5]. Following the completion of a PhD study in Austria, notable advancements were made in 1994. The EU and Austrian governments authorized the commercial production of CLT in 1998. The use of CLT construction started to grow in Canada and the US in the 2000s after it gained popularity in Europe. Additional details regarding the history of the CLT can be found in the literature [6-9]. The DLT is not new technology, just like the CLT [10-13]. Diagonalization was used with dowels, veneer, plywood, and nails. In 2011, the term DLT was coined in the literature by Bejtka and Bosil, who studied it under in-plane loads [14, 15].

Bending stiffness is used in engineering calculations in a variety of formulas that are specified for the CLT by international design specifications, such as the Swedish, American, Canadian, European, Italian, and Croatian ones. To our knowledge, no specification has a comparable formula for the bending stiffness of the DLT. Nonetheless, some suggestions regarding the stiffness of the DLT can be found online. The DLT's missing formula is developed and presented in this paper. In addition to bonding with finite properties and the ply angle, it also covers the CLT and hybrid species for the first time.

The available CLT stiffness formulas are determined by cross-sectional and material elastic properties. First, for the DLT's flexural stiffness, some websites suggested combining the Hankinson formula with the shear analogy method. It is an inaccurate idea [16]. Rather than transverse behavior, the Hankinson formula was empirically introduced from the idea of in-plane failure theory.

Arnold utilized Hankinson's formula for the off-axis modulus of elasticity transformation in his PhD thesis [17]. Arnold's method was adopted in some DLT studies under transverse load. In addition to being incorrect, that adoption lacks any references that link the in-plane behavior of the Hankinson formula to the Bernoulli-Euler theory, which emphasizes the flexural behavior of panels.

Second, the length-to-width ratio of a classic CLT is between 6 and 30, according to the American Plywood Association [18]. This ratios range demonstrates how well CLT panels perform in engineering calculations as beams. Under in-plane loads rather than transverse ones, DLT has been studied in the literature [17, 19, 20].

Finally, the "stability of wood components" has multiple interpretations. It might refer to dimensional stability, which is outside the scope of this applied engineering paper and is a scientific subject. To our knowledge, there isn't a single study on the structural stability of DLT panels in the literature [21-23]. A few sources, nevertheless, have looked into the stability coefficient for buckling of CLT under axial compression.

2. Analytical Model and Solution

Though DLT was invented decades ago and to our knowledge, there isn't a rigorous analytical model with

solution for the CLT and DLT in the literature. This study has filled the gaps. It produced an exact series-type solution that accounts for real bonding stiffness, interlayer slip, and ply angle. The model and solution are applicable to both DLT and CLT.

For the analysis and design of CLT and DLT, perfect rigid bonding is widely assumed in the literature [24, 25]. This is not a realistic assumption. The unrealistic formulation is replaced in this paper by the constitutive differential equations of the bending moment curvature and the interlayer slip. After considering the newly introduced ply angle and non-rigid bonding between the layers, the governing differential equations are derived, and then the boundary conditions are considered. Since there are no analytical solutions in literature that incorporate the bonding with finite stiffness, Fourier trigonometric series are used to solve and satisfy all of the governing equations and boundary conditions. The series-type solutions are used in the literature because they can adapt to different loading scenarios and boundary conditions [26-28].

The validity and accuracy of our formulation and solution were ascertained by comparing the predicted serviceability measure, i.e., deflections, with selected specific analytical results obtained in the literature by credible sources. All the outcomes are in good-satisfactory agreement. Parametric analysis was performed to determine the effects of bonding stiffness, interlayer slip, and ply angles on the structural performance.

It is worth noting that the rigorous foundation for the type of non-rigid bonding engineered laminates has been reported in the literature by this author [29-40]. Our contributions were acknowledged by NASA to be among the top 1.5 % of pertinent achievements.

For our analytic model and solution, consider a panel of length and width L and b, and depth d1 and d2, where subscript 1 refers to the upper layer. The panel is subjected to a uniform transverse load of intensity q. Assuming the layers have the same curvature, then:

$$\frac{M_1}{E_1 I_1} = \frac{M_2}{E_2 I_2} = -\frac{d^2 w}{d x^2}$$
(1)

Eq. (1) can be rewritten as follows:

$$M_1 + M_2 = -EI \frac{d^2 w}{d x^2},$$
 (2)

where M_i is the bending moment, E_i is the modulus of elasticity, I_i is the moment of inertia, w is a deflection, and x is the coordinate axis with the origin at the panel's left end. The panel is not under external in-plane load, thus:

$$\sum_{i=1}^{2} N_i = 0, (3)$$

where $N_{i}\xspace$ is the internal axial force. At the interface between the layers, the compatibility of deformations

must be satisfied. The compatibility condition is expressed as [27, 29, 44]:

$$V = k du = k (u_1 \cos^2 \theta - u_2),$$
 (4)

where V is the interlayer shear flux, u_i is the in-plane displacement, θ is the ply angle, and k is the bonding stiffness. The equilibrium equation of the upper layer requires that:

$$E_1 A_1 \cos^2 \theta \frac{d^2 u_1}{dx^2} + V = 0,$$
 (5)

where A is a cross-sectional area. The applied bending moment, M_a , and the resisting moment, M_r , must be equal. Thus

$$M_a = -EI \frac{d^2w}{dx^2} - E_1 A_1 \cos^2\theta d \frac{du_1}{dx}, \qquad (6)$$

where d is the center-to-center distance. According to the fundamentals of conventional engineering mechanics [27, 43, 44],

$$\frac{d^2M}{dx^2} = EI \frac{d^4w}{dx^4} - E_1 A_1 \cos^2\theta \, d\frac{d^3u_1}{dx^3} = p, \qquad (7)$$

where p is the applied load, and $EI = E_1 I_1 \cos^2 \theta - E_2 I_2$.

The solution to the governing equations must satisfy the boundary conditions, which for simply supported panels are w=0 at x=0 and x=L, and N at x=0 and x=L.

The displacements and applied load are expressed as a trigonometric series as follows:

$$w(x) = \sum_{i=1,2,\dots}^{\infty} W_i \sin \alpha_i x , \qquad (8)$$

$$u_1(x) = \sum_{i=1,2,..}^{\infty} u_{1i} \cos \alpha_i x$$
, (9)

$$u_{2}(x) = \sum_{i=1,2,..}^{\infty} u_{2i} \cos \alpha_{i} x , \qquad (10)$$

$$p = \sum_{i=1,2\dots}^{\infty} p_i \sin \alpha_i x, \qquad (11)$$

in which $\alpha_i = \frac{i \pi}{L}$ and $p_i = \frac{4 p_o}{i \pi}$ for uniformly applied load of intensity p_o , $p_i = \frac{2 P \sin \alpha_i x_p}{L}$ for concentrated load P at distance x_p , and $p_i = \frac{4 M_o}{i \pi}$ for end moment M_o . The series equations satisfy the boundary conditions. The p_i of various loads can be added for a combined effect.

The first governing equation is obtained by substituting Eqs. (9) and (10) in Eq. (3). Thus

$$E_1 A_1 u_{1i} \cos^2 \theta + E_1 A_1 u_{2i} = 0$$
 (12)

Substituting Eqs. (8), (9), and (10) in Eq. (5) results in

The second governing equation is obtained from Eq. (12) in conjunction with Eq. (13), as follows:

$$\frac{k d \alpha_i W_i}{E_1 A_1} - \left(\alpha_i^2 + \frac{k \cos^2 \theta}{EA}\right) u_{1i} = 0, \qquad (14)$$

in which $EA = \frac{E_1 A_1 E_2 A_2 \cos^2 \theta}{E_1 A_1 \cos^2 \theta - E_2 A_2}$. Now, by combining Eqs. (11), (8), and (9) with

Eq. (7), the third governing equation is obtained as

$$\operatorname{EI} \alpha_i^4 W_i + \operatorname{E}_1 \operatorname{A}_1 \operatorname{d} \, \cos^2 \theta \, \alpha_i^3 \, u_{1i} = p_i \qquad (15)$$

Finally, W_i is found by solving Eqs. (14) and (15), thus

$$W_{i} = \frac{p_{i}}{\alpha_{i}^{4} \left(\text{EI} - \frac{k \, d^{2}}{\alpha_{i}^{2} + \frac{k}{FA}} \right)} \tag{16}$$

By following the previous analytical formulation using three layers panel, the solution is found as follows for uniform load of intensity po:

$$W_{i} = \frac{4 p_{o} L^{2}}{(\pi i)^{5} EI} \left[\frac{2 d^{2}}{\left(2 d^{2} - \frac{EI}{E_{1} A_{1} \cos^{2}\theta} + \frac{1}{k} \left(\frac{i \pi}{L}\right)^{2}\right)} - 1 \right], \quad (17)$$

where 1, 2, and 3 refer to the upper, middle, and bottom layers, respectively, and d is the distance between mid-plane of adjacent layers.

3. Validation

The literature incorporates numerical and experimental studies rather than a comprehensive formulation and solution for DLT and CLT panels. Thus, the above model and solutions were validated using existing specific special cases with existing solutions obtained in credible sources in the literature [25, 27, 43]. For this purpose, we considered two- and three-layer panels with the following properties: span L=3.66 m, width=0.24 m, thickness=0.3 m, E=10 MPa, and a uniform load of q=575 kN/m. The following table compares the results for two- and three-layers panel:

Table 1 shows that the solutions are in satisfactory good agreement.

Table 1. Comparison of the maximum deflection of a simple CLT panel.

Number of Layers	Present Sol.	Ref. [25, 27, 43]	Col. 2/Col.3, %
2	0.242	0.249	97
3	0.219	0.231	95

4. Practical Method – Hussein's Formula

In order to determine the bending stiffness of mechanically connected CLT panels, formulas have

been provided by the Croatian, American, Canadian, European, Italian, and Swedish standards [8, 25, 47-50]. The rolling shear is used in those formulas to represent the shear deformation. This paper's author created a number of rigorous solutions for engineered laminates with bonding that had finite stiffness, and NASA recognized his contribution as one of the top 1.5 % of accomplishments [29-40].

Based on our completed pertinent studies, a novel practical formula is coined in this section for CLT and DLT with imperfect bonding. According to classical Euler-Bernoulli theory, the effective bending stiffness of CLT panels with perfect bonding, EIeff, is as follows:

$$(EI)_{eff} = \sum_{i=1}^{n} (EI)_i, \qquad (18)$$

where I is the moment of inertia, E is the modulus of elasticity, and i is the ith layer. By following the exact procedure in our previous completed studies, the following equation is obtained for the apparent stiffness.

$$(EI)_{app} = \sum_{i=1}^{n} (EI \cos^2 \theta)_i + \frac{(EA a^2)_i}{\left[1 + \frac{\pi^2 EA}{K L^2}\right]_i}, \quad (19)$$

where a is the distance of the centroid to the panel's centroid, K is the real bonding stiffness, and L is the span of the panel. Because literature has no studies to compare with, we used the solutions of selected particular cases where solutions were obtained by credible scholars to validate the above formula. This approach is known in literature where data is lacking [26, 27, 28, 51]. Like Hankinson's formula, Eq. (19) is referred to as the Hussein's formula. Both formulas are for practical use, easy to understand, readily applicable to real design situations, concise, and use readily available variables.

Table 2. Comparison of the bending stiffness	using
Eq. (19) and other sources.	

Ply angle	classic/Eq. (19), %
0/0	97.2 %
0/90/0	97.0 %
90/0/90	100 %
0/0/0	96.7 %
90/0/0/90	100 %
0/90/0/90/0	100 %
0/0/0/0/0	100 %
0/0/0/0/0/0/0	100 %
90/0/90/0/90/0/90	100 %

Equation (19) was used to calculate the stiffness of ten CLT and DLT with different configurations. All of the above panels have $E=10^6$ psi. All the results are in good satisfactory agreement.

5. Discussion

Fig. 1 illustrates how the ply angle affects the deflection for both two- and three-layer panels. When the angle increases from 10 degrees by two orders of magnitude, the deflection increases by 14 %; however, when the angle increases from 60 degrees by the same order, the deflection increases by 164 %. This finding suggests that the DLT is not suitable for applications under transverse loads. The CLT has the least deflection due to high bending stiffness. The deflection at any ply angle can now be computed analytically using the series solution that has been presented. There is no literature on such a tool.



Fig. 1. Effect of ply angle on DLT deflection.

The load-slip curves were obtained in some studies [44, 46]. The combined effects of bonding stiffness and ply angle on the maximum deflections of two- and three-layer panels are shown in Fig. 2. The deflection is more sensitive to changes in the bonding stiffness, k, value in its lower range for all ply angles. After stiffness reaches a certain value, the bonding could be regarded as practically rigid. Using a common engineering sense, a very rigid bonding would be unnecessary with weak species, and the opposite would be unwise. This discovery is not mentioned in literature.



Fig. 2. Effect of bonding stiffness on deflection.

Fig. 3 shows the combined effects of ply angle and bonding stiffness on interlayer slippage. It is seen that slippage essentially becomes insensitive to bonding stiffness after undergoing a sharp nonlinear change at low k values. This is another significant finding. For instance, when the ply angle was changed from 30 to 60 degrees with k=0.5 MPa, the slippage increased by 97%; however, when the angle was changed from zero degrees by the same order of magnitude, the slippage increased by only 18%.



Fig. 3. Effect of bonding stiffness and ply angle on interlayer slippage.

Figs. 2 and 3 provide an answer to what constitutes perfect bonding. That corresponds to the k value at which the change in deflection is considered practically insensitive.

Fig. 4 depicts the effect of ply angle on bending stiffness. Again, as the ply angle increases the stiffness decreases. This is another proof that CLT performs better than DLT under transverse loads. The figure demonstrates that the bending stiffness remains relatively high up to an angle of about 40 degrees.



Fig. 4. Effect of ply angle on bending stiffness of 3-layer DLT.

6. AI Toolkit for CLT and DLT

In general, the goal of AI is to use computers to carry out tasks that call for human intelligence. The reader can find the wide breadth and deep depth of the realm of AI in several noticeable review articles [52-55]. The connection between CLT and DLT and AI is noticeably absent in literature; thus, there is no pertinent AI-based research and literature. This paper is an attempt to close this gap. Our AI conceptualization is based on the fundamentals of expert systems and algorithms, that are subsets of the AI technologies, that were applied in diverse areas [56-67]. Because we applied engineering mechanics, which is a deterministic field, the toolkit is also deterministic. Thus, statistical simulations or database sets are not required.

Figs. 5 and 6 show the general outline and the dashboard of the AI toolkit. The toolkit was developed based on known available ES studies [46, 68-75].

Typical input data includes geometrical dimensions, materials' properties, and the applied load. Once the input data is entered, the toolkit executes If-Then loops to calculate the deflection using the developed analytical model explained previously. The expert examines the output and does modifications and changes, then re-runs the toolkit.



Fig. 5. General outline of the AI toolkit.

	AI TOOLKIT DASHBOA	RD
Upper width, m	Upper MOE, Pa	Span, m
Middle width, m	Middle MOE, Pa	Applied load, N/m
Lower width, m	Lower MOE, Pa	Ply angle, degrees
Upper depth, m	Upper bonding stiffness, N/m/m	Deflection, m
Middle depth, m	Lower bonding stiffness, N/m/m	
Lower depth, m		

Fig. 6. Dashboard of the AI toolkit.

Fig. 7 illustrates the iterative search process for the design that meets the requirements. An expert and the toolkit can exchange diagnostic iterative cycles of input and output that end when both of the two sides are satisfied.



Fig. 7. An AI-based design process.

The MS Excel Solver can be used to modify the data based on defined constraints by the expert. Fig. 8 shows a typical snapshot of the Solver. A basic understanding of programming can be used to find a solution without using the app in Fig. 8.

et Objective:		1. Objective function				
Ta: (Max	OMB	O Value Of:	2. Limi	2. Limiting value	
(by Changin	g Variable	Cells				
3. Cha	nging	variables			1	
Sybject to I	he Consta	aints:				
4 Constraints				≜dd		
4.00	Qhange Delete Beset A0				Change	
					Delete	
					Eeset All	
				*	Load/Save	
🖬 Mage U	nconstrain	ned Variables No	on-Negative			
Sglect a Sol Method:	ving (GRG Nonlinear				
Solving M	ethod					
Select the Simplex er problems	GRG Noni Igine for I that are n	linear engine fo inear Solver Pro on-smooth.	e Solver Problems that blems, and select the	are smooth noni Evolutionary engl	inear. Select the LP ne for Solver	

Fig. 8. Excel Solver to reach satisfactory solution.

The suggested methodology enables both human and virtual brains to jointly arrive at an acceptable design. The real human brain is preserved throughout the process because engineering design takes into account incalculable factors like practical considerations, experience, interdisciplinary collaborations, heuristic rules, etc.

7. Conclusions

Artificial intelligence became a potent tool for innovative engineering solutions and design in today's world. Nonetheless, literature has no AI applications on engineered DLT and CLT. This paper closes this gap using ES, which is one of the AI subsets. The paper encompasses two interconnected sections.

In the first section, the paper critiques the evidently misrepresented literature on DLT. Also, a novel comprehensive model for the DLT and CLT was developed. The rigorous engineering mechanics-based governing equations are derived, then solved using Fourier trigonometric series. The series-type approach was essential because a closed-form solution would be unfeasible and is adaptable for various loading scenarios and boundary conditions. For the first time, essential properties such as the bonding stiffness, the interlayer slips, and the ply angle are considered in the formulation. Several examples are used to validate and verify the formulation. A parametric analysis is conducted to ascertain how these properties impact structural performance. The findings suggest that the smaller the ply angle, the higher the bending stiffness, and the bonding stiffness must be considered in the design when serviceability is the quantity of interest. The paper addressed what really constitutes perfectly rigid bonding that dominates the literature. In addition, a practical formula was developed for bending stiffness. It includes bonding stiffness, ply angle, and cross-sectional properties. It was also validated and verified. The formula is practical, easy to understand, readily applicable to real design situations, concise, uses readily available variables, and allows for calculations without unnecessary complexity.

In the second section, a novel AI-based methodology and toolkit are coined. The real human expert and virtual intelligence $\frac{1}{100}$ cooperate in achieving satisfactory designs. An expert and the toolkit can exchange diagnostic iterative cycles of input and output that end when both of the two sides are satisfied. The Excel MS Solver is introduced to facilitate the implementation of the human-computer exchange. The suggested methods enable both human experts and virtual brains to jointly arrive at a satisfactory design. The real, actual human brain is preserved throughout the process because engineering design takes into account incalculable factors like practical considerations, experience, interdisciplinary collaborations, heuristic rules, etc.

References

- [1]. Plywood and Engineered Wood, https://www.apawood.org/apas-history
- [2]. History of CLT, https://sites.cnr.ncsu.edu/clt-panels/ history-of-cross-laminated-timber/
- [3]. Dowel Laminated Timber, https://structurecraft.com/ materials/mass-timber/dlt-dowel-laminated-timber
- [4]. 100 Projects UK CLT, https://www.thinkwood.com/wp-content/uploads/ 2019/08/Think-Wood-Publication-100-Projects-UK-CLT.pdf
- [5]. Cross-Laminated Timber, https://en.wikipedia.org/ wiki/Cross-laminated timber
- [6]. H. Ren, et al., A state-of-the-art review on connection systems, rolling shear performance, and sustainability assessment of cross-laminated timber, *Engineering Structures Journal*, Vol. 317, 2024, 118552.
- [7]. M. Mohammad,, et al., Introduction to cross laminated timber, *Wood Design Focus*, Vol. 22, Issue 2, 2021, pp. 3-12.
- [8]. M. Jeleč, D. Varevac, V. Rajčić, Cross-laminated timber (CLT) – a state of the art report, *Journal of the Croatian Association of Civil Engineers*, Vol. 70, Issue 2, 2018, pp. 75-95.
- [9]. Cross-laminated timber (CLT) turns 100, https://www.archpaper.com/2023/08/cross-laminatedtimber-clt-turns-100/
- [10]. M. Lechner, P. Dietsch, S. Winter, Veneer-reinforced timber, in *Proceeding of the World Conference on Timber Engineering (WCTE'20)*, Santiaco, 24-27 August 2020.
- [11]. Method for Manufacturing Diagonal Plywood, https://patents.google.com/patent/US9527222B2/en
- [12]. Dowel Laminated Timber, https://structurecraft.com/ materials/mass-timber/dlt-dowel-laminated-timber.
- [13]. Brettstapel, https://en.wikipedia.org/wiki/Brettstapel

- [14]. I. Bejtka, Cross (CLT) and Diagonal (DLT) Laminated Timber as Innovative Material for Beam Elements, *KIT Scientific Publishing*, 2011.
- [15]. R. Bosl, Zum Nachweis des Trag- und Verformungsverhaltens von Wandscheiben aus Brettlagenholz, Universität der Bundeswehr München, 2002.
- [16]. N. Mascia, E. Nicolas, R. Todeschini, Comparison between Tsai-Wu failure criterion and Hankinson's Formula for tension, *Wood. Research*, Vol. 56, Issue 4, 2011, pp. 499-510.
- [17]. M. Arnold, Mechanical properties of Diagonal Laminated Timber (DLT) with respect to pointsupported mass timber slabs, PhD Thesis, *Technische Universität München*, 2023.
- [18]. Cross Laminated Timber (CLT), https://www.apawood.org/cross-laminated-timber
- [19]. J. Turesson, Diagonal compression of Cross-Laminated Timber, MD Thesis, *Lulea University of Technology*, 2016.
- [20]. D. Buck, O. Hagman, Production and in-plane compression mechanics of alternatively angled layered Cross-Laminated Timber, *Journal of BioResources*, Vol. 13, Issue 2, 2018, pp. 4029-4045.
- [21]. M. Sciomenta, et al., Buckling analyses of Cross Laminated Timber panels, in *Proceedings of the World Conference on Timber Engineering*, Oslo, Norway, 19-22 June 2023, pp. 2761-2767.
- [22]. Q. Ye, et al., Analysis and calculation of stability coefficients of Cross-Laminated Timber axial compression member, *Polymers*, Vol. 13, Issue 23, 2021, 4267.
- [23]. G. Wu, et al., The effect of the bearing width on the buckling capacity of partially loaded CLT member, *Buildings*, Vol. 12, Issue 1, 2022, 84.
- [24]. A. Noor, C. Bert, Computational Models for Sandwich Panels and Shells, *American Society of Mechanical Engineers*, 1996.
- [25]. E. Karacabeyli, B. Douglas, CLT Handbook, *FPInnovations*, 2013.
- [26]. Y. Zhang, L. Zhang, S. Zhang, Exact series solutions of composite beams with rotationally restrained boundary conditions: static analysis, *Archive of Applied Mechanics*, Vol. 92, 2022, pp. 3999-4015.
- [27]. R. Szilard, Theories and Applications of Plate Analysis: Classical, Numerical and Engineering Methods, *John Wiley & Sons*, 2004.
- [28]. S. Timoshenko, J. Goodier, Theory of Elasticity, McGraw-Hill Company, 1951.
- [29]. R. Hussein, Thermal Stress in Flat Metal-Faced Sandwich Panel, *Concordia University*, Canada, 1978.
- [30]. R. Hussein, Structural Behavior of Sandwich Panels, Concordia University, 1980.
- [31]. K. Ha, R. Hussein, P. Fazio, Analytic solution for continuous sandwich plates, *Journal of the ASCE Engineering Mechanics Division*, Vol. 108, Issue 2, 1982, pp. 228-241.
- [32]. P. Fazio, R. Hussein, K. Ha, Sandwich beam-columns with interlayer slips, *Journal of Engineering Mechanics*, Vol. 108, Issue 2, 1982, pp. 354-366.
- [33]. R. Hussein, Sandwich plates with interlayer slips, *Journal of Engineering Mechanics*, Vol. 110, Issue 4, 1984, pp. 493-506.
- [34]. R. Hussein, Orthotropic sandwich plates with interlayer slip and under edgewise loads, *Journal of Structural Engineering and Mechanics*, Vol. 17, Issue 2, 2004, pp. 153-166.
- [35]. R. Hussein, P. Fazio, Thermal nonlinear behavior of sandwich panels, *Journal of Experimental Mechanics*, Vol. 25, 1985, pp. 140-144.
- [36]. R. Hussein, K. Ha, P. Fazio, Thermal stresses in sandwich panels with interlayer slips, *Journal of Thermal Stresses*, Vol. 12, Issue 2, 1989, pp. 191-207.
- [37]. R. Hussein, K. Ha, P. Fazio, Thermal stresses in sandwich plates, *Journal of Thermal Stresses*, Vol. 12, Issue 3, 1989, pp. 333-349.
- [38]. R. Hussein, P. Fazio, K. Ha, Analytical evaluations of local failures in sandwich panels, *Building and Environment Journal*, Vol. 26, Issue 2, 1991, pp. 209-215.
- [39]. R. Hussein, P. Fazio, K. Ha, Effects of bonding stiffness on thermal stresses in sandwich panels, *Journal of Aerospace Engineering*, Vol. 5, Issue 4, 1992, pp. 480-490.
- [40]. R. Hussein, Application of plate theory to laminated wood composites with non-rigid adhesive, *International Wood Products Journal*, Vol. 1, Issue 1, 2010, pp. 35-42.
- [41]. J. Goodman, E. Popov, Layered wood systems with interlayer slip, *Journal of the Structural Division*, Vol. 94, Issue 11, 1968, pp. 2535-2548.
- [42]. E. Amana, et al., Theoretical and experimental studies of nailed and glued stressed-skin components, *Journal* of Wood Science, Vol. 4, Issue 19, 1967.
- [43]. S. Timoshenko, D. Young, Theory of Structures, McGraw-Hill Company, 1965.
- [44]. A. Haftkhani, H. Hematabadi, Effect of layer arrangement on bending strength of Cross-laminated (CLT) manufactured from Poplar, *Buildings*, Vol. 12, Issue 5, 2022, 608.
- [45]. H. Li, et al., An experimental and modeling study on apparent bending moduli of cross-laminated bamboo and timber (CLBT) in orthogonal strength directions, *Journal Case Studies in Construction Materials*, Vol. 16, 2022, e00874.
- [46]. E. Nilsson, Characterization of Cross Laminated Timber properties, MD Thesis, *Lund University*, 2021.
- [47]. E. Karacabeyli, S. Gagnon, Canadian CLT Handbook, *FPInnovations*, 2019.
- [48]. European Committee for Standardization, EN 1995-1-1: Eurocode 5 Design of Timber Structures, *The European Union*, 2004.
- [49]. Technical Recommendations for Construction, *Italian* National Research Council, 2006.
- [50]. A. Gustafsson, The CLT Handbook Facts and Planning, RISE Research Institutes of Sweden, 2019.
- [51]. S. Timoshenko, S. Woinowsky-Krieger, Theory of Plates and Shells, *McGraw-Hill Book Company*, 1959.
- [52]. D. Onatayo, et al., Generative AI applications in architecture, engineering, and construction: trends, implications for practice, education & imperatives for upskilling – a review, *Architecture Journal*, Vol. 4, Issue 4, 2024, pp. 877-902.
- [53]. R. Rajput, C. Singh, Review of artificial intelligence and human thinking, *Turkish Journal of Computer and Mathematics Education*, Vol. 12, Issue 1, 2021, pp. 855-860.
- [54]. I. Cinar, Y. Taspinar, M. Koklu, Artificial Intelligence Applications in Engineering, *ISRES Publishing*, 2021, pp. 107-125.
- [55]. V. Hulwane, Review of AI in civil engineering, ACEE Annals of Civil and Environmental Engineering, Vol. 8, Issue 1, 2024, pp. 48-51.

- [56]. M. Maher, Expert Systems for Civil Engineers: Technology and Applications, *ASCE*, 1987.
- [57]. C. Krishnamoorthy, S. Rajeev, Artificial Intelligence and Expert Systems for Engineers, CRC Press, 1996.
- [58]. R. Hussein, Knowledge-based tools for monitoring and management, and design of the engineered infrastructure construction systems, in Advances in Computers and Software Engineering Review, Vol. 2, *IFSA Publishing*, 2019, pp. 199-250.
- [59]. R. Hussein, AI-based tools for performance and monitoring of sustainable built and natural environments, and the climate, in Advances in Artificial Intelligence: Reviews, Vol. 1, *IFSA Publishing*, 2019, pp. 185-202.
- [60]. R. Hussein, Treatise on sustainable infrastructure construction: green composites, cross laminated/mass timber, wood truss connectors, nondestructive technologies, health assessment and monitoring: utility poles and geofoam, in Advances and Technologies in Building Construction and Structural Analysis, *IntechOpen Publisher*, 2021, pp. 1-44.
- [61]. R. Hussein, What a wasted municipal solid waste's data: an innovative data analytics driven strategic framework for future MSW decision making, in *Proceedings of the International Conference on Advances in Signal Processing and Artificial Intelligence (ASPAI'22)*, Corfu, Greece, 19-21 October 2022, pp. 86-88.
- [62]. R. Hussein, A Coined knowledge-based computational toolkit for biomasses-based sustainable infrastructure: Python, VB, MATLAB, in *Proceedings of the 5th International Conference on Advances in Signal Processing and Artificial Intelligence (ASPAI'23)*, Tenerife, Spain, 7-9 June 2023, pp. 288-290.
- [63]. R. Hussein, What does the MSW industry need: data technologies or more waste?, in *Proceedings of the 5th International Conference on Advances in Signal Processing and Artificial Intelligence (ASPAI'23)*, Tenerife, Spain, 7-9 June 2023, pp. 258-260.
- [64]. R. Hussein, Vibration-based nondestructive assessment of rotational stiffness of structural connections, *International Journal on Recent and Innovation Trends in Computing and Communication*, Vol. 3, Issue 3, 2015, pp. 1696-1698.
- [65]. R. Hussein, Smart technologies for health assessment and monitoring of infrastructure components, *International Journal on Recent and Innovation Trends in Computing and Communication*, Vol. 3, Issue 4, 2015, pp. 2456-2460.
- [66]. R. Hussein, Applications of lignin-based foam as a load-bearing component in engineered laminates, *International Wood Products Journal*, Vol. 15, Issue 2, 2024, pp. 128-138.
- [67]. R. Hussein, Computer toolkit for the structural analysis and design of cross laminated loadbearing components, US TXu 2-245-722 2020, USA, 2020.
- [68]. S. Latifi, Modelling and testing of CLT panels for evaluation of stiffness, MD Thesis, *Linnaeus University*, 2021.
- [69]. M. Maher, Expert systems for structural design, *Journal of Computing in Civil Engineering*, Vol. 1, Issue 4, 1987, pp. 270-283.
- [70]. M. Maher, HI-RISE and beyond: directions for expert systems in design, *Computer-Aided Design Journal*, Vol. 17, Issue 9, 1985, pp. 420-427.
- [71]. J. Giarratano, G. Riley, Expert Systems: Principles and Programming, *Thompson Course Technology*, 1998.

- [72]. G. D'Aronco, An algorithm for numerical modelling of Cross-Laminated Timber structures, *MD Thesis*, Università degli Studi di Padova, 2015.
- [73]. M. Zecchetto, Una procedura numerica per il progetto di edifici in X-Lam, MD Thesis, *Università degli Studi di Padova*, 2015.
- [74]. C. Johansson, E. Johansson, Modeling of Cross Laminated Timber in FE analysis, PhD Thesis, *Chalmers University of Technology*, 2021.
- [75]. E. Gezer, et al., Determining the optimum layer combination for Cross-Laminated Timber panels according to timber strength classes using artificial neural networks, *Journal of BioResources*, Vol. 19, Issue 3, 2024, pp. 4899-4917.

(017)

Removing EOG Artifacts from EEG Recordings Using Deep Learning

C. O'Reilly 1-4 and S. Huberty 5,6

 ¹ Department of Computer Science and Engineering, University of South Carolina, Columbia, SC, USA
 ² Institute for Mind and Brain, University of South Carolina, Columbia, SC, USA
 ³ Artificial Intelligence Institute, University of South Carolina, Columbia, SC, USA
 ⁴ Carolina Autism and Neurodevelopment Research Center, University of South Carolina, Columbia, SC, USA
 ⁵ Department of Pediatrics and Neurology, University of Southern California, Los Angeles, CA, USA
 ⁶ Division of Neurology, Children's Hospital Los Angeles, Los Angeles, CA, USA E-mail: christian.oreilly@sc.edu

Summary: The electroencephalogram (EEG) directly measures the electrical activity generated by the brain. Unfortunately, it is often contaminated by various artifacts, notably those caused by eye movements and blinks (EOG artifacts). Such artifacts are usually removed using an independent component analysis (ICA) or other blind source separation techniques. However, it is difficult to assess whether subtracting EOG components estimated through ICA removes some neurogenic activity. It is crucial to address this question to avoid biasing EEG analyses. Toward that objective, we developed a deep learning model for EOG artifact removal that exploits information about eye movements available through eye-tracking (ET). Using a multimodal EEG and ET open-access dataset, we trained within-subject a long short-term memory (LSTM) model to predict the component of EEG signals predictable from ET data. We further used this ET-informed evaluation of EOG artifacts to investigate the sensitivity and specificity of ICA. Our analysis indicates that although ICA is very sensitive to EOG, it has a comparatively low specificity. These results motivate further research on EEG artifact removal to develop approaches with higher EOG rejection specificity.

Keywords: Electroencephalogram eye tracking, Deep learning, Independent component analysis, Electrooculogram, LSTM.

1. Introduction

Electroencephalography (EEG) is a non-invasive neuroimaging technique used to record the electrical activity generated by the brain. EOG (electrooculogram) artifacts in EEG recordings refer to the electrical signals generated by eve movements and eye blinks due to the potential difference between the cornea and the retina, which acts as an electric dipole. When the eyes move, their dipoles also move, generating a change in the field of electric currents propagating instantaneously (quasistatic approximation [1]) through the volume of the head and reaching the EEG electrodes [2]. Thus, EOG artifacts are omnipresent in EEG signals. They usually account for much of the EEG variance, particularly in the channels near the eye. These artifacts are particularly problematic in EEG tasks requiring eye movements, as they can obscure the neural activity related to the experimental task. Therefore, developing techniques for removing EOG artifacts with high specificity is critical for EEG research, particularly for analyses of frontal connectivity involving non-lagged homotopic synchronization, which cannot be reliably distinguished from instantaneous electrical volume conduction [3].

Various techniques have been proposed to separate and remove the artifactual components from the neural components of the EEG signals. Among that class of algorithms, Independent Component Analysis (ICA) has become arguably the most widely adopted approach [4] and is used in most EEG preprocessing pipelines [5, 6]. However, although this approach has the benefit of retaining the complete EEG time course, it may fail to remove all the artifacts (insufficient sensitivity) or may distort the neural signals (insufficient specificity). Since ICA is an unsupervised algorithm, confirming that the components labeled as artifacts do not include neural signals is challenging, and some neural signals can inadvertently and unknowingly be lost when these components are subtracted from the EEG data.

While ICA can be used to identify and remove various artifacts (EOG, ECG, power line noise), alternative artifact reduction techniques for removing EOG artifacts specifically exist. For example, EOG channels from electrodes placed near the eyes can be used in a linear regression to estimate the corresponding EOG artifact in EEG channels, which can then be subtracted from the signal [7, 8]. However, EOG electrodes differ from EEG electrodes only in their placement near the eyes. Thus, they also pick up neural signals or other types of artifacts (e.g., muscle contraction, sweating), which diminishes their utility as a reference signal for EOG activity.

In this study, we aim to use deep learning to remove EOG artifacts in an EEG/eye-tracking (ET) dataset and compare the performance of this approach to ICA, the most prominent approach for EOG removal.

2. Methods

2.1. Datasets

To develop and test the proposed model, we leveraged the EEGEyeNet dataset [9]. This dataset contains recordings from 356 healthy adults, including simultaneously collected high-density 129-channel EEG data synchronized with video-infrared ET. The ET data include two channels for the position in X and Y and one for the pupil size. Both raw and preprocessed data are available in the EEGEveNet dataset. The experiment contains three tasks: a pro- and antisaccade task, a visual symbol search task, and the Large Grid task. For our study, we used the latter [10]. For this task, 30 participants were asked to look at dots appearing at 25 different positions, distributed across the whole surface of a screen (see Fig. 1). Each dot is presented for 1.5 to 1.8 seconds, in а pseudo-randomized order (see [9] for details on this pseudo-randomization). The central dot was displayed three times, while the other dots were presented once per block. Each of the six runs contained five blocks, totaling 810 stimuli per participant. Each run is saved as a separate recording, providing 177 recordings (i.e., three were missing).



Fig. 1. Positions of presented dots during the Large Grid task (blue/red circles) overlayed with the gaze distribution.

2.2. Outlier Rejection

Before running analyses, we excluded recordings with noisy or unreliable ET data, which could have been due to many reasons, the most likely being poor eye-tracker calibration. We identified outliers by first computing the mean squared difference between the event-related response (ERR) of every recording and the ERR averaged across recordings

$$e = \overline{(X - \overline{X}^r)^2}^t, \tag{1}$$

where X is a matrix of x and y gaze coordinates, and the bars with r and t represent the averaging across the recording and time dimensions, respectively. We computed these errors for each ET channel, dot stimuli, and recording. We then computed the 25t^h, 50th, and 75th quantiles of the distribution of these errors. To detect outliers in individual recordings, we used the classic outlier rejection formula, $e > Q_{50} + k(Q_{75} - Q_{50})$ with k = 6, and rewrote

$$\frac{e - Q_{50}}{Q_{75} - Q_{50}} > k \tag{2}$$

to average the left-hand term across channels and dots (i.e., event types) before comparing these averages against the threshold. Using this outlier criterion, we excluded eight recordings from further analyses. Most of them were for different runs from the same participant.

We confirmed participant compliance with the Large Grid Task instructions and the quality of the ET data for the remaining recordings by displaying the two-dimensional kernel density estimation of the distribution of the X/Y pixel coordinates for every dot in the large grid (Fig. 1). For this computation, we determined the gaze position as the average position in the $t \in [0.3, 1.0]$ s time window, with t = 0 being the stimulus presentation onset.

2.3. Preprocessing

Minimally and *maximally* preprocessed versions of the EEGEyeNet dataset are available [11]. These two alternative preprocessing are defined by the *Automagic* toolbox [12]. We used the minimally preprocessed version, which includes bad channel detection and interpolation and EEG filtering to the 0.5-40 Hz band. This minimal preprocessing does not include ICA artifact rejection since this step would remove the EOG artifacts necessary for our study. The authors of EEGEyeNet synchronized the EEG and ET signals and confirmed the absence of synchronization errors exceeding 2 ms.

To make our analysis more computationally efficient, we filtered to the 1-30 Hz band before downsampling the signals to 100 Hz using MNE-Python [13]. Although the dataset is recorded with a sampling rate of 500 Hz, EOG signals are limited to relatively low frequencies due to natural biomechanical constraints imposed on the kinematics of eye movements. Thus, such a high sampling rate significantly increases the network size (i.e., multiply by a factor of five the long short-term memory (LSTM) input shape and the associated weights to learn) without adding relevant information.

For machine learning, recordings were epoched into contiguous 1 s segments (this should not be confused with the concept of training epochs in deep learning), resulting in a 3D matrix of size $n_{segments} \times n_{channels} \times n_{times}$. We also set an average reference. For our comparison with ICA, we used the Extended Infomax approach [14] as implemented in MNE-Python. EOG-associated components were detected automatically as those labeled as *eye blink* by MNE-ICALabel [15], which ports to Python the functionalities of ICLabel [16]. ICLabel has six additional classes of independent components: *brain*, *muscle artifact*, *heartbeat*, *line noise*, *channel noise*, and *other*.

2.4. Deep Learning Model

The overarching idea of our approach is to train a recurrent neural network (RNN) to predict EEG signals only from ET signals. Of course, only a small portion of the EEG signals will be predictable from ET signals, and this predictable portion will be due to EOG artifacts and potentially some neural and non-neural correlates of eye movements (e.g., electromyographic signals due to the activation of the muscle required for eye movements). More formally, for the EEG signal matrix Y (EEG channels \times time) and the ET matrix X (ET channels \times time), we model this relationship as Y = f(X) + R where R is a residual matrix containing the neural signals and potentially non-eye-movementrelated artifacts, and f is a nonlinear function we want to learn by adjusting the RNN weights to minimize the mean square amplitude of R. For this task, we used a 2-layer LSTM with three features corresponding to ET channels and 64 hidden states whose outputs get pruned with a 0.5 dropout layer. A final fully connected layer maps the LSTM internal states to 129 outputs corresponding to the EEG channels (Fig. 2). We implemented this model in PyTorch and fitted it using the ADAM optimizer with a 0.01 learning rate and a mean-square-error (MSE) loss function. We found that 1000 training epochs (not to be confused with the EEG 1 s epochs) were enough to reach a stable training loss. We did not attempt to test for generalizability across participants or recordings. Rather, we wanted to test if the mapping between ET signals and their impact on EEG was learnable within participants. Thus, we implemented no hold-out or cross-validation. That is, the mapping was learned independently for all 177 recordings, and we used each of these mappings to clean the EEG of the corresponding recording only.

2.5. Analysis

EOG signals are known to affect mostly frontal EEG channels. To validate the capability of the neural network to detect and remove EOG noise, we computed the percentage of signal removed per channel as

$$\Delta = 1 - \frac{R}{Y} \tag{3}$$

We also defined the reaction time (RT) to a stimulus as the moment the gaze position changed by 5% of maximal response amplitude. Using this RT, we used an approach similar to the one adopted in [17] to

characterize the sensitivity and specificity of removing eye artifacts based on ET versus automated ICA. In this approach, we defined a pre-reaction time (pre-RT) segment where we expect no noise, here defined as the window from the start of the baseline period (-0.2 s) to the reaction time (RT), and a post-RT window where EOG artifacts are expected (RT to 1 s). Defining the signal (S) as the root mean square (RMS) amplitude of the original recording (Y) and the noise (N) as the RMS amplitude of what has been removed by the artifact removal approach (i.e., N = Y - R), we can define the classic measure of signal-to-noise ratio (SNR) in dB as

$$SNR = 10 \cdot \log_{10}\left(\frac{S}{N}\right) \tag{4}$$

A large SNR before the reaction time is indicative of a high specificity (i.e., clean signals do not get distorted by artifact removal), and a low SNR after the reaction time is indicative of a good sensitivity (i.e., more noise has been detected and removed by the cleaning approach).



Fig. 2. The deep neural network we designed for EOG/EEG decounfounding. a) Block diagram and corresponding equations for the LSTM model used in the deep neural network. In the equations, \odot stands for the Hadamard product, σ is the sigmoid function, lowercase letters are vectors, and uppercase letters are matrices. The matrices *W* and vectors *b* are learned during the training, b) Deep neural network architecture using two LSTM layers and one fully connected layer.

3. Results

A demonstration of cleaned EEG signals using the proposed model and ICA compared to the original EEG signals is shown in Fig. 3.

Further, evoked responses to gazes at dots confirm that the model accurately removes the EOG contamination in the EEG recordings (Fig. 4).

We also looked at the distribution of the predicted noise over the scalp (Fig. 5), averaged across participants. This distribution clearly shows the bias toward frontal regions, with about 70 % of the amplitude of the recorded signals in prefrontal and frontal channels being due to EOG artifacts. Further, the proposed model predicted less EOG in the pre-RT period than ICA. The fixation of the participants' gaze during the period preceding the apparition of a stimulus (see the bottom right panel of Fig. 4 for an illustration of this) suggests that the proposed model distorts the EEG signals less than ICA.



Fig. 3. Top panel: EEG signals for a typical recording before (gray) and after EOG artifact reduction using the proposed model (green) and ICA (blue). Bottom panel: X and Y gaze positions (reported in pixel coordinates) during the same period as the top panel.



Fig. 4. The evoked response to dot 25 before and after EOG removal (top panel) and the average eye movement (X/Y pixels) during those trials (bottom panel). The vertical red dashed line represents the average RT across participants, corresponding to the moment they began shifting their gaze toward the dot displayed at the beginning of the trial (i.e., at t = 0 s).

Next, we used the evoked EEG time series (averaged for each dot) before and after EOG artifact reduction to test the sensitivity and specificity of the proposed deep learning model and ICA. In Fig. 4, we can observe that the deep learning model seems more conservative in EOG reduction than ICA and distorts signals less in the pre-RT period. To quantify the specificity and sensitivity of the proposed model for reducing EOG artifacts, we computed the SNR as described in (4). The averaged SNR values for the preRT (specificity) and post-RT (sensitivity) periods within subjects (across runs) and computed a paired ttest to compare the SNR values for the proposed model versus ICA. Fig. 6a illustrates an example of this approach for channel E25 and dot 25. For this combination of channel and dot, the average values for the Model and ICA indicate a higher specificity for the model but a higher sensitivity for ICA. We repeated this process for all channels and dots and aggregated the result on topomaps for specificity (Fig. 6b) and sensitivity (Fig. 6c). The results suggest that this tendency toward higher specificity but lower sensitivity for the model compared to ICA can be generalized across the scalp, except for a higher sensitivity of the model in the posterior region of the scalp.



Fig. 5. Topographic plots showing the spatial projection of the predicted EOG signal (across participants), as predicted by the model (top row) and by ICA (bottom row). The spatial projection is shown for the pre-RT period (left column) and the post-RT period (right column). We used only trials for dot for this figure.



Fig. 6. Performance of our approach based on eye-tracking signals (ET) versus ICA, a) Example of analysis of sensitivity and specificity for a specific channel (E25, a frontal channel) and dot 25, b) Comparative performance of the specificity for the two approaches across the scalp. For each channel and dot, specificity is determined as illustrated in panel a. SNR values are then averaged within subjects (across runs), and a paired t-test is computed to compare the SNR values for ET vs ICA. We counted +1 when the specificity was larger for ET than ICA, -1 when it was smaller, and 0 when it was not statistically different (p > 0.05). The sum of these scores is computed across the 27 dot conditions and displayed as a topomap, c) Same as for b, but for sensitivity.

4. Discussion

This study demonstrated a novel approach for EOG artifact rejection in EEG signals recorded simultaneously with ET. The code implementing this available GitHub approach is on (https://github.com/lina-usc/eog-learn). The emergence and popularization of such recordings [18] has opened new possibilities by providing reference signals closely associated with EOG generation. One of the constant challenges of EOG rejection is the absence of a ground truth for evaluating the effectiveness of proposed approaches. This common situation limits the options available to investigators to 1) using synthesized recordings with a known ground truth but questionable face and ecological validity or 2) using recorded EEG with unknown ground truth. The addition of ET data, although not providing ground truth for EOG artifacts per se, partly mitigates this thorny issue by providing reliable information on eye motions usable for inferring EOG artifacts.

Leveraging these signals, we adopted a *data-driven* black-box approach to EOG artifact removal from EEG signals. The *data-driven* qualifier in this statement comes from using ET signals and deep learning to empirically map the association between the movement of the eyes and its impact on EEG signals. We use the *black-box* qualifier to contrast with an approach using ET signals and a physiological model of how eye movements generate EOG artifacts. Our approach does not consider any knowledge specific to the application at hand. The solution relies on the generic task of learning an arbitrary relationship between an input and an output given enough data. This task can be addressed by deep neural networks, which have been shown to work as universal function approximators [19]. Because of the adoption of a generic solution, conceptually, this approach may be suitable for other applications (e.g., removal of electrocardiogram artifacts in EEG, correction for the effect of motion on electrocardiogram signals), as long as the source of the contamination is due to a process for which we have a separate reference signal.

Crucially, the approach we adopted provides an opportunity to assess the performances of blind source separation conducted with ICA more objectively. This technique currently dominates the field [4]. It has been shown to perform well in general, and our analyses support this position in many ways. However, ICA for EOG rejection is known to have limitations [17]. More importantly, although its capacity to remove EOG artifacts can be readily evaluated on noisy EEG signals (e.g., see Fig. 3), the degree to which it may distort neural signals is more difficult to establish. For example, ICA tends to distort the phase of EEG signals [20, 21], a key element in neural dynamics and a property essential for all functional connectivity metrics based on phase consistency (e.g., coherence, phase locking value, phase lag index). Our experiment demonstrated that eye movement information can be used to remove EOG artifacts effectively while

distorting neural signals significantly less. Although the lack of a ground truth creates some uncertainty in the interpretation of the SNR-based measures of sensitivity and specificity, our approach has shown a much higher specificity than ICA. In general, ICA has shown a higher sensitivity. However, since we do not have the ground truth for the effect of eye movements on the EEG, we cannot rule out the possibility that the apparent superior sensitivity of ICA in post-RT windows could also be partly due to an over-correction of ICA.

Interestingly, our approach was more sensitive to parts of the scalp that typically are less impacted by EOG artifacts (i.e., central/occipital regions; see Fig. 6c). Our comparison with ICA relied on the automated classification of independent components associated with eye movement artifacts. The classifier we adopted (i.e., ICLabel) has been designed using machine learning and a large dataset of independent components annotated by experts. This classification is, therefore, vulnerable to biases associated with our understanding of the topological appearance of EOG artifacts. EOG artifacts are known to have the most impact on the frontal region. However, although electrical dipoles generated close to the forehead may have their strongest effect on that part of the scalp, their field wraps around virtually the whole head (with decreasing amplitude due to attenuation). Our results suggest that independent components selected for rejection may tend to undercorrect for these more distant effects. This observation highlights the most significant contribution of this approach: using ET to assess the impact of eye movement on EEG may provide us with a more reliable and objective assessment of EOG artifact topography. We can then use this assessment to develop new methods or correct existing methods that do not require the availability of ET data.

5. Limitations

The need for synchronized EEG and ET recordings constitutes the most obvious limitation of the approach we proposed in this paper. However, although we may use this approach directly to clean EEG in such datasets, this application was not the main reason for this study. We would argue that using such a dataset to develop methods that can highlight the limitations of current techniques and suggest possible ways to remedy these shortcomings without requiring ET data is of greater interest. We focused our comparisons on the automated ICA approach because of its popularity for EOG removal.

For our specific application, we attempted to learn the predictable part of the EEG based on ET information. This part is arguably a minor portion of the EEG makeup. The learning task is, therefore, complicated by the relatively low percentage of predictable information in the signals. We demonstrated that even with a relatively small amount of data (i.e., single recordings), it is possible to learn that relationship with a satisfying degree of precision. However, we could possibly improve the sensitivity and specificity of EOG removal by using a larger training set. The degree to which the performance is saturated with the current size of the training data is unknown.

The data used for this study (i.e., looking at many predefined targets on a screen) were particularly well-suited for our analysis and for learning the mapping between eye movement and EOG artifacts. However, we used information on the structure of the task only for performance analysis. Our training did not rely explicitly on the properties of the experimental protocol (i.e., the whole recording was segmented in 1s epochs and passed to the training routine without any information on the stimuli). However, implicit characteristics (e.g., the systematic coverage of the whole screen area by eye movements) may have been beneficial.

It is also worth considering that the relationship learned between eye movement and the predictable part of the EEG is made up of various contributions, including a component due to the artifact generated by the movement of the eye (the component we generally want to remove) and the neural activity systematically correlated with the eye movement, such as the neural signals controlling the movement of the eye and the neural activity created by the change in visual stimuli when the line of sight shifted. Analysts may or may not want to remove these latter components depending on the hypotheses under investigation. The approach considered in this study cannot disentangle these different components. However, it offers a powerful framework for investigating these components, for example, by using virtual reality to experimentally control changes in the visual field as a function of eye movement.

Lastly, we decided in this study to perform learning and testing on single recordings. One advantage of this approach is that it is constrained to the recordings themselves (e.g., it is independent of the subject sample size). This approach is, therefore, tailored to the specificity of the participants (e.g., the specific shape and electrical properties of the head of the participant affect how electrical currents generated by the movement of the eyes travel and are recorded by the EEG system) and of the recording (i.e., effects of the experimental protocol, environment factors, etc.). Thus, our study did not aim for the generalizability and reusability of these deep learning models. Future work could consider training a single model across a large sample to target generalizability and reusability. It is also possible that by benefiting from a larger sample for training, the mapping learned would be more precise. Whether the gain obtained from training across subjects would offset the loss of precision due to interindividual and within-individual/betweenrecordings variability is currently unknown. Alternatively, pretrained models may provide an advantageous middle ground.

6. Conclusion and Future Works

We presented a deep-learning approach that leverages ET information in synchronized EEG/ET recordings to remove EOG artifacts from EEG recordings. Most importantly, we demonstrated how to harness this objective source of information to benchmark existing approaches and better understand their limitations. In future works, we plan to address the limitations associated with this black-box approach by designing a generative model of how EOG artifacts are generated from eye movements relying on physiological knowledge. Provided that 1) we dispose of enough information to develop a faithful generative model of EOG artifact from measurements of the eves position and 2) given that precise ET information is available, this source of artifact could be removed accurately from the EEG. Furthermore, in combination with the approach we presented here, the EOG and the neural component associated with eye movements could then be disentangled, allowing more precise analyses of neural activity.

References

- R. Plonsey, D. B. Heppner, Considerations of quasistationarity in electrophysiological systems, *Bull. Math. Biophys.*, Vol. 29, Issue 4, Dec. 1967, pp. 657-664.
- [2]. O. Dimigen, Optimizing the ICA-based removal of ocular EEG artifacts from free viewing experiments, *NeuroImage*, Vol. 207, Feb. 2020, 116117.
- [3]. C. O'Reilly, M. Elsabbagh, Intracranial recordings reveal ubiquitous in-phase and in-antiphase functional connectivity between homotopic brain regions in humans, *J. Neurosci. Res.*, Vol. 99, Issue 3, Mar. 2021, pp. 887-897.
- [4]. X. Jiang, G.-B. Bian, Z. Tian, Removal of artifacts from EEG signals: a review, *Sensors*, Vol. 19, Issue 5, Feb. 2019, 987.
- [5]. L. J. Gabard-Durnam, A. S. Mendez Leal, C. L. Wilkinson, A. R. Levin, The Harvard Automated Processing Pipeline for Electroencephalography (HAPPE): standardized processing software for developmental and high-artifact data, *Front. Neurosci.*, Vol. 12, 2018, 97.
- [6]. S. Huberty, J. Desjardins, T. Collins, M. Elsabbagh, C. O'Reilly, PyLossless: A non-destructive EEG processing pipeline, *bioRxiv*, 2024, 2024-01.
- [7]. G. Gratton, M. G. H. Coles, E. Donchin, A new method for off-line removal of ocular artifact, *Electroencephalogr. Clin. Neurophysiol.*, Vol. 55, Issue 4, Apr. 1983, pp. 468-484.
- [8]. R. J. Croft, R. J. Barry, Removal of ocular artifact from the EEG: a review, *Neurophysiol. Clin. Clin. Neurophysiol.*, Vol. 30, Issue 1, Feb. 2000, pp. 5-19.
- [9]. M. B. Plomecka, A. Kastrati, N. Langer, EEGEyeNet, Dataset, OSF, May 2021, https://osf.io/ktv7m/
- [10]. J. Son, et al., Evaluating fMRI-based estimation of eye gaze during naturalistic viewing, Cereb. Cortex N. Y. N 1991, Vol. 30, Issue 3, Mar. 2020, pp. 1171-1184.
- [11]. A. Kastrati, *et al.*, EEGEyeNet: a simultaneous electroencephalography and eye-tracking dataset and

benchmark for eye movement prediction, *arXiv* preprint, Nov. 10, 2021, arXiv:2111.05100.

- [12]. A. Pedroni, A. Bahreini, N. Langer, Automagic: Standardized preprocessing of big EEG data, *NeuroImage*, Vol. 200, Oct. 2019, pp. 460-473.
- [13]. A. Gramfort *et al.*, MNE software for processing MEG and EEG data, *NeuroImage*, Vol. 86, Feb. 2014, pp. 446-460.
- [14]. T.-W. Lee, M. Girolami, T. J. Sejnowski, Independent component analysis using an extended Infomax algorithm for mixed subGaussian and superGaussian sources, *Neural Comput.*, Vol. 11, Issue 2, Feb. 1999, pp. 417-441.
- [15]. A. Li, J. Feitelberg, A. P. Saini, R. Höchenberger, and M. Scheltienne, MNE-ICALabel: Automatically annotating ICA components with ICLabel in Python, J. Open Source Softw., Vol. 7, Issue 76, 2022, 4484.
- [16]. L. Pion-Tonachini, K. Kreutz-Delgado, S. Makeig, ICLabel: An automated electroencephalographic independent component classifier, dataset, and website, *NeuroImage*, Vol. 198, Sep. 2019, pp. 181-197.

- [17]. D. Srishyla, S. J. Webb, M. Elsabbagh, C. O'Reilly, B. Team, Eye-movement artifact correction in infant EEG: A systematic comparison between ICA and Artifact Blocking, J. Neurosci. Methods, 2025, (in press).
- [18]. O. Dimigen, B. V. Ehinger, Regression-based analysis of combined EEG and eye-tracking data: Theory and applications, J. Vis., Vol. 21, Issue 1, Jan. 2021, 3.
- [19]. K. Hornik, M. Stinchcombe, H. White, Multilayer feedforward networks are universal approximators, *Neural Netw.*, Vol. 2, Issue 5, Jan. 1989, pp. 359-366.
- [20]. R. Montefusco-Siegmund, P. E. Maldonado, C. Devia, Effects of ocular artifact removal through ICA decomposition on EEG phase, in *Proceedings of the 6th International IEEE/EMBS Conference on Neural Engineering (NER'13)*, Nov. 2013, pp. 1374-1377.
- [21]. R. W. Thatcher, E. P. Soler, D. M. North, G. Otte, Independent components analysis 'artifact correction' distorts EEG phase in artifact free segments, *J Neurol. Neurobiol.*, Vol. 6, Issue 4, 2020, pp. 5-7.

(018)

GNSS Non-Line-of-Sight (NLOS) Error Repairing in Challenging Urban Environments with Channel Attention and Inception-based Deep Learning Network

Zhiqiang Wang^{1,2}, Ni Zhu³ and Ruiwen He¹

 ¹ De Vinci Higher Education, De Vinci Research Center, Paris, France
 ² Nantes Université, École Centrale de Nantes, CNRS, LS2N, UMR 6004, F-44000 Nantes, France
 ³ AME-GEOLOC, Univ. Gustave Eiffel, F-44344 Bouguenais, France E-mail: ni.zhu@univ-eiffel.fr

Summary: Global Navigation Satellite Systems (GNSS) provide absolute positioning for a wide range of applications. However, their performance can be severely degraded due to multipath and Non-Line-of-Sight (NLOS) receptions caused by surrounding environments, particularly in urban canyons. The traditional statistic-based mitigation methods encounter bottlenecks since the errors from these local effects are difficult to model accurately using existing distributions. To handle this issue, this paper proposes a deep-learning framework to estimate the additional distances of the range measurements caused by NLOS receptions. The customized architecture involves four main modules: a non-linear data transformation to rescale the data, a channel attention mechanism to weigh different features, a generative Convolutional Neural Network (CNN) network to augment the feature map, and an inception module to enhance the feature extraction from a multi-hierarchical level. The model was trained and tested with real urban GNSS data, which shows promising results, achieving an RMSE of 6.76 m and a 75 % prediction error of 0.71 m, demonstrating higher accuracy than the current state-of-the-art.

Keywords: GNSS, NLOS, Error modeling, Deep learning.

1. Introduction

Global Navigation Satellite Systems (GNSS) are widely used across various sectors, from entertainment to safety and reliability critical applications. GNSS positioning algorithms rely on measuring the signal Time-of-Arrival (ToA) between the satellite and the receiver to determine the range and use trilateration to estimate the user's position. However, GNSS range accuracy is often degraded in urban environments due to multipath and Non-Line-of-Sight (NLOS) caused by signal reflection and diffraction from surrounding obstacles. This can lead to huge positioning errors which is harmful for many critical location-based applications.

Traditional methods detect and exclude faulty GNSS measurements based on consistency checks, which have limitations in harsh urban scenarios. On the one hand, when the majority of the satellite measurements are erroneous, which is frequent in urban, healthy satellites could be wrongly excluded. On the other hand, excluding satellites can worsen the already limited satellite visibility in urban areas, leading to poor Dilution of Precision (DOP) and further degrading GNSS performance.

That is why, another key research branch focuses on fault detection and weighting instead of exclusion. De-weighting faulty measurements can effectively reduce the impact of errors in filter propagation while maintaining optimal satellite geometry. However, the basic signal Quality Indicators (QI), such as Signal to Noise Ratio (SNR) and satellite elevation, become less effective in urban scenarios since they do not fully capture observation quality. Our goal is to go beyond current limitations by repairing GNSS NLOS errors. By estimating and removing the additional distance caused by NLOS reception, the erroneous measurements can be repaired and used efficiently. The objective of this paper is to design a customized deep learning-based framework to estimate the additional range caused by NLOS. The main contributions of this paper are as follows:

- Propose a deep learning architecture leveraging channel attention to weight features, a Generative Convolutional Neural Network (GCNN) network to augment and enrich the feature map, as well as an inception module to enhance the feature extraction from a multi-hierarchical level;
- 2) Train and evaluate the model performance on real data collected in urban scenarios.

The remainder of the paper is organized as follows: Section 2 presents the related state-of-the-art research. The proposed methodology is presented and evaluated respectively in Section 3 and Section 4. Section 5 draws conclusions and future work.

2. State-of-the-art

GNSS multipath and NLOS error modeling approaches can be classified into two categories: model-based and data-driven. Traditional statistical model-based approaches focus on bounding measurement errors with known distributions modulated by certain signal QIs. The most common ones are based on Gaussian distribution while inflating the variance as a function of signal-in-space ranging errors (SISRE), C/N0 [1-2], satellite elevation [3], pseudorange residuals or code-minus-carrier observables (CMC) [4]. Better performances can be obtained by hybridizing multiple indicators including LOS/NLOS information from the map data, as shown in [5]. However, these models require proper calibration according to the specific receiver and environment. New OIs are considered leveraging the GNSS pseudorange residual from different time slots [6] or the probability of NLOS reception predicted by machine learning models [7]. Besides, some Bayesian approaches are proposed in the literature to model the GNSS multipath error. [8] leveraged Dirichlet Process Mixture (DPM) to model the GNSS measurement errors, extending the problem to non-Gaussian and nonlinear situations. However, the implementation complexity as well as the hyperparameter optimization for the DPM become non-negligible issues.

Recently, more and more data-driven approaches have emerged to address GNSS multipath and NLOS receptions in stringent urban environments. The majority of them are designed for signal classifications using Support Vector Machine (SVM) [9], CNN [10], Gradient Boosting, LSTM [11] and their ensembling [12] are implemented to classify method LOS/NLOS/multipath. The features are extracted from different levels, from the correlator to the raw measurement level. [13] provided a review for ML-based GNSS multipath mitigation and implemented a Fully Connected Neural Network (FCNN) to benchmark on open access data. The overall classification accuracy varies from 70 % -98 % depending on the testing scenarios and data size.

Besides, some research focuses on AI-based multipath or NLOS error regression. [14] proposed a

CNN-based method by converting correlator outputs into images to estimate multipath parameters. Tests on synthetic data show satisfactory results for attenuation coefficient, code delay and difference in Doppler shift but the estimation of phase difference still needs to improve. [15] proposed an LSTM network to estimate directly the weights allocated to each satellite showing improvement in positioning accuracy using real data. [13] made a benchmark using FCNN to predict the pseudorange error with a final RMSE of 15.14 m. The challenges remain on how to properly design the architecture of the AI model to fully leverage the potential of the features and to make the model robust and generalizable for different cities.

3. Methodology

The proposed deep-learning framework is composed of four main modules: data non-linear transformation. channel attention. feature augmentation, and an inception-based deep regressor. The global architecture is illustrated in Fig. 1. The following seven basic GNSS features are used as input: pseudorange, carrier phase, Doppler, SNR, satellite elevation. azimuth, epoch-wise Normalized Pseudorange Residual (NPR). The NPR can be written as follows:

$$NPR = \frac{PR_i - PR_{min}}{PR_{max} - PR_{min}},\tag{1}$$

where PR_i is pseudorange residual at epoch *i*, PR_{max} and PR_{min} represent respectively the maximum and the minimum pesudorange residuals at each epoch.



Fig. 1. Architecture of the proposed channel attention and inception-based CNN network.

3.1. Non-linear Quantile Transformation

A quantile transformation is applied for data preprocessing to make different features comparable and to ensure the model learns the intrinsic patterns of the data instead of their ranges. Here, the original values of the features are mapped to a Gaussian distribution while the outliers are mapped to the boundaries of the distribution.

3.2. Channel Attention

Channel attention is originally designed for image processing, especially in CNN to capture the most important features across different channels [16]. Here we have adapted the channel attention to our scenario as shown in Fig. 2, to allocate different weights to features while enhancing the network's ability to learn complex patterns. First, a Global Average Pooling (GAP) and a Global Max Pooling (GMP) are applied to obtain the global information of the input feature vector with the dimension of 1×7 . The outputs of the pooling layers, i.e., F_{avg} and F_{max} , are concatenated together then pass through a shared Multi-layer Perception (MLP) and a sigmoid activation function. Finally, the attention weight vector is obtained and multiplied by the original features.

7th International Conference on Advances in Signal Processing and Artificial Intelligence (ASPAI' 2025), 8-10 April 2025, Innsbruck, Austria



Fig. 2. Channel attention mechanism applied on GNSS single channel feature map.

3.3. Feature Augmentation

Considering the sensor noises, potential perturbations on the feature values as well as the limited number of features, we propose here to use a generative CNN-based network to augment the input feature. As shown in Fig. 1, the feature augmentation module is composed of several convolutional and upsampling operations layers, which gradually augment the 1D feature vector into 2D feature map. A leaky Rectified Linear Unit (LeakyReLU) activation function is applied at the end of each layer, which is calculated as follows:

$$LeakyReLU(x) = \begin{cases} x \text{ if } x > 0\\ \alpha x \text{ otherwise} \end{cases}$$
(2)

where α is a small constant representing the slope for negative input, here $\alpha = 0.01$. Compared to the traditional ReLU activation function, LeakyReLU provides a small, non-zero gradient for negative inputs to prevent "dying neurons" where the gradient becomes zero and the neuron stops learning.

At the end of the feature augmentation module, the original 1D feature vector with size 1×7 is augmented into 2D feature map whose size is 28×28 .

3.4. Deep Inception-based Regressor

The reconstructed feature map is then fed into an inception-based regressor. As shown in Fig. 3, the proposed inception architecture is composed of convolution layers with different depths. The structure is inspired by [17] but is different from it because the final output is the sum of the outputs from different convolution paths instead of the concatenation. The operation allows feature extraction from different hierarchical levels to enhance the robustness and the generalization ability of the model.

4. Performance Evaluation

4.1. Experimental Setup

The dataset was collected with a GNSS receiver (Ublox LEA 6T) mounted on the roof of an

experimental vehicle in the urban area of Nantes, France. The ranges of additional distances of the NLOS signals are labeled by a ray-tracing technique using highly accurate ground truth and city 3D models. The dataset has a total of 17903 epochs with 130073 observations, which is split into 80 % for training and 20 % for testing.



Fig. 3. Proposed inception module for regression, which fosters the multi-scale feature extraction.

Prediction Accuracy Assessment

The performance is evaluated based on prediction error, which is defined as follows:

$$PredError_i = y_i - \hat{y}_i, \tag{3}$$

where y_i represents the true value of the range additional distance at the epoch *i*, \hat{y}_i represents the predicted range additional distance using one AI model.

The model is trained in two modes: one using all signals, including both LOS and NLOS, and the other using only NLOS signals, assuming an upstream signal classifier is present so only the additional range distances from NLOS are trained and predicted.

The following assessment criteria are used: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), the coefficient of determination R^2 (the closer to 1, the better) as well as the 75 % and 95 % of the absolute error. The proposed framework is benchmarked with the Extreme Gradient Boosting (XGB) optimized by the Optuna optimizer as well as a simple Multi-Layer Perceptron (MLP). The implemented MLP is composed of two Fully Connected (FC) layers, both with the same dimensions: an input size of 7 and an output size of 150.

Fig. 4 shows the CDF comparison of the absolute prediction error with the benchmark methods and proposed method respectively for the whole test data and the NLOS-only test data. Table 1 summarizes the corresponding statistics. The proposed method achieves a global better regression performance than the XGB and MLP, with an RMSE of 6.76 m for LOS / NLOS together and 8.10 m for NLOS-only data.



Fig. 4. CDF of absolute prediction error for the whole test data and NLOS-only test data using XGB and the proposed method.

4.3. Ablation Study

To further evaluate the effectiveness of the proposed architecture, an ablation study is done particularly on the impact of feature selection as well as the attention mechanism.

The following alternate feature sets are considered for comparison with the proposed 7-feature set:

- A most commonly used 3-feature set from state-of-the-art research [13, 18] including pseudorange residual, SNR and satellite elevation;
- A 5-feature set including SNR, NPR, CMC, LT and the number of satellites.

Table 2 summarizes the performance statistics using different feature sets, where the proposed 7-feature set achieves the best performance.

Data	Method	RMSE [m]	MAE [m]	AE 50 % [m]	AE 75 % [m]	AE 95 % [m]	R2 [m]
	XGB	8.72	2.63	0.06	1.85	11.26	0.66
LOS & NLOS	MLP	7.99	3.74	1.91	5.10	16.35	0.51
	Proposed	6.76	1.74	0.01	0.71	8.94	0.74
	XGB	10.05	3.48	2.40	5.95	23.11	0.55
NLOS-only	MLP	12.37	7.14	4.41	8.91	25.37	0.48
	Proposed	8.10	2.53	1.12	3.65	17.78	0.65

Table 1. Model Prediction Error Comparison on Test Data.

 Table 2. Prediction error comparison using different feature sets.

Data	Feature set	RMSE [m]	MAE [m]
	3-feature set	10.33	3.31
LOS & NLOS	5-feature set	10.29	3.43
	Proposed 7- feature set	6.76	1.74
	3-feature set	14.57	6.08
NLOS only	5-feature set	14.46	6.77
	Proposed 7- feature set	8.10	2.53

To study the impact of the attention mechanism, we compare the model performance with and without this module. Table 3 presents a summary of the prediction performance with and without the attention mechanism. It shows that excluding the attention mechanism results in an approximate increase of 0.2 m in overall RMSE and MAE for both LOS & NLOS data as well as NLOS-only data.

These ablation studies highlight the effectiveness of the proposed methodology, particularly in feature set selection and architecture design.

5. Conclusion and Perspectives

This paper proposed a customized deep learning architecture that is designed based on the characteristics of the GNSS data to predict the error caused by local effects such as NLOS and multipath. The network includes a channel attention module to highlight important features, a generative CNN module to further augment existing features, and an inception-based regressor to make the NLOS error estimation. In this way, only using seven most basic features, the network is able to regress the range error accurately. Testing results on real data collected in urban environments show that the RMSE of the proposed architecture can achieve 6.76 m and a 75 % absolute error of 0.71 m while these metrics using XGB are respectively 8.72 m and 1.85 m, and using MLP are respectively 7.99 m and 5.10 m.

Fable 3. Ablation	study of the	attention	mechanism.
-------------------	--------------	-----------	------------

Data	Configuration	RMSE [m]	MAE [m]
	With Attention Mechanism	6.76	1.74
LOS & NLOS	Without Attention Mechanism	6.92	1.97
	With Attention Mechanism	8.10	2.53
NLOS only	Without Attention Mechanism	8.37	2.61

Future work remains on testing and improving the robustness and the generalization ability of the model on different city configurations.

References

- Z. Zhang, B. Li, Y. Gao, Y. Shen, Real-time carrier phase multipath detection based on dual-frequency C/N0 data, *GPS Solutions*, Vol. 23, 2019, pp. 1-13.
- [2]. P. R. Strode, P. D. Groves, GNSS multipath detection using three-frequency signal-to-noise measurements, *GPS Solutions*, Vol. 20, 2016, pp. 399-412.
- [3]. H. Hartinger, F. K. Brunner, Variances of GPS phase observations: the sigma- model, *GPS Solutions*, Vol. 2, 1999, pp. 35-43.
- [4]. S. Khanafseh, B. Kujur, M. Joerger, T. Walter, S. Pullen, J. Blanch, K. Doherty, L. Norman, L. de Groot, B. Pervan, GNSS multipath error modeling for automotive applications, in *Proceedings of the 31st International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+'18)*, 2018, pp. 1573-1589.
- [5]. N. Zhu, GNSS propagation channel modeling in constrained environments: Contribution to the improvement of the geolocation service quality, PhD Thesis, Université de Lille, 2018.
- [6]. J. Liu, B.-g. Cai, D.-b. Lu, J. Wang, A local weighting method for GNSS receiver autonomous integrity monitoring using pseudorange residuals, in *Proceedings of the International Conference on Intelligent Transportation Systems (ITSC'18)*, 2018, pp. 3067-3074.
- [7]. L. Li, M. Elhajj, Y. Feng, W. Y. Ochieng, Machine learning-based GNSS signal classification and weighting scheme design in the built environment: a comparative experiment, *Satellite Navigation*, Vol. 4, Issue 1, 2023, 12.
- [8]. A. Rabaoui, N. Viandier, E. Duflos, J. Marais, P. Vanheeghe, Dirichlet process mixtures for density estimation in dynamic nonlinear modeling: Application to GPS positioning in urban canyons, *IEEE Transactions on Signal Processing*, Vol. 60, Issue 4, 2011, pp. 1638-1655.
- [9]. T. Suzuki, Y. Amano, NLOS multipath classification of GNSS signal correlation output using machine learning, *Sensors*, Vol. 21, Issue 7, 2021, 2503.

- [10]. C. Jiang, Y. Chen, B. Xu, J. Jia, H. Sun, Z. He, T. Wang, J. Hyyppä, Convolutional neural networks based GNSS signal classification using correlator-level measurements, *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. 46, 2022, pp. 61-66.
- [11]. N. Zhu, C. Belemoualem, V. Renaudin, A portfolio of machine learning-based GNSS LOS/NLOS classification in urban environments, in *Proceedings of the IEEE SENSORS*, 2023, pp. 1-4.
- [12]. S. J. Koiloth, D. S. Achanta, P. R. Koppireddi, Ml-based LOS/NLOS/multipath signal classifiers for GNSS in simulated multipath environment, *Aerospace Systems*, Vol. 7, 2023, pp. 237-250.
- [13]. P. Xu, G. Zhang, B. Yang, L.-T. Hsu, Machine learning in GNSS multipath/NLOS mitigation: Review and benchmark, *IEEE Aerospace and Electronic Systems Magazine*, Vol. 39, Issue 9, 2024, pp. 26-44.
- [14]. T. Gonzalez, A. Blais, N. Couellan, C. Ruiz, Multipath parameters estimation in physically based synthetic environment using robust deep neural regression, in *Proceedings of the International Technical Meeting of The Institute of Navigation*, 2023, pp. 808-822.
- [15]. I. Sbeity, C. Villien, B. Denis, E. V. Belmega, Lstm-based GNSS localization using satellite measurement features jointly with pseudorange residuals, *Sensors*, Vol. 24, Issue 3, 2024, 833.
- [16]. Y. Xue, J. Qin, Partial connection based on channel attention for differentiable neural architecture search, *IEEE Transactions on Industrial Informatics*, Vol. 19, Issue: 5, 2023, pp. 6804-6813.
- [17]. R. He, H. Benhabiles, F. Windal, G. Even, C. Audebert, A. Decherf, D. Collard, A. Taleb-Ahmed, A CNN-based methodology for cow heat analysis from endoscopic images, *Applied Intelligence*, Vol. 52, 2022, pp. 8372-8385.
- [18]. R. Sun, G. Wang, W. Zhang, et al., A gradient boosting decision tree based GPS signal reception classification algorithm, *Applied Soft Computing*, Vol. 86, 2020, 105942.

(019)

Diagnosing Plant Leaf Disease with THz Sensor and Digital Signal Processing

Janez Trontelj, Andrej Švigelj and Janez ml. Trontelj

University of Ljubljana, Faculty of Electrical Engineering, Tržaška 25, 1000 Ljubljana, Slovenia Tel.: + 3864768333 e-mail: janez.trontelj1@guest.arnes.si

Summary: This paper proposes a system for early plant leaf disease detection using highly sensitive THz sensors and digital signal processing. Pathogens, such as fungi, bacteria, viruses, and environmental stressors, usually cause infections of the plants and damage their leaves. They strongly influence plant health, yield, and overall aesthetic value. They are usually directly correlated with changes in the dielectric constant of leaf tissues and moisture, which is known to have a strong absorption coefficient for THz radiation. This phenomenon leverages THz radiation's unique, non-invasive interaction with water molecules and plant tissues. We developed and manufactured our low-cost, highly sensitive THz sensors. The article describes some examples of practical usage of these sensors in biology. We will show that the amount of leaf water content or internal leaf moisture for different leaf areas can be easily detected with a THz camera and displayed as a grayscale image using some digital signal processing. Combining optical and THz images of the leaf can mitigate the risk of infections and promote overall plant health.

Keywords: Leaf disease, THz rays, THz sensor, Spectral signature in THz range, THz camera.

1. Introduction

This paper proposes a system for early plant leaf disease detection using several highly sensitive THz sensors assembled in a portable THz camera working at room temperature. The sensor array uses digital signal processing (DSP) to display a THz image of the leaf in almost real-time as a grayscale image. Light gray represents dry areas, and darker gray represents water-soaked areas.

This approach is novel since, for the time being, THz technology is not widely used in agricultural diagnostics. Most existing plant disease detection methods rely on optical, infrared imaging, or hyperspectral analysis. The latter operates in the visible to near-infrared spectrum, while THz waves can penetrate certain materials and offer insights into subsurface features like detecting moisture reach areas. Hyperspectral imaging is surface-sensitive and relies only on emitted or reflected light from the surfaces.

Types of leaf diseases are generally divided into four typical classes: physiological disorders, fungal, bacterial, and viral diseases. Internal leaf moisture plays a critical role in developing many leaf diseases. On the other hand, diseases and stress cause typical structural and biochemical changes in leaf tissues. These changes usually produce distinct spectral signatures in the THz detection range.

THz waves are electromagnetic waves between microwaves and infrared light in the frequency range of 0.1–10 THz. They are sensitive to water content and can penetrate non-conductive materials, including plant tissues [1, 2], making them ideal for assessing internal moisture and detecting several moisture-related anomalies in plant leaves.

2. Materials and Methods

We developed and manufactured our highly sensitive, low-cost THz sensor [3-6]. It consists of a nano bolometer coupled with an antenna. We experimented with several antenna types of different sizes, mostly dipole and wideband antennas. We also constructed a prototype of a THz camera with an array of sensors for simultaneously detecting two different THz frequencies. An example of a sensor array with a nano bolometer and wideband type antenna is shown in Fig. 1.



Fig. 1. Array of THz sensors and nano bolometer sensor coupled with wide band antenna.

Symptoms of leaf diseases include discoloration, spots or blotches, rusts, deformation, fungal growth, yellowing or browning edges, frost damage, mottled patterns, water-soaked lesions, curling, etc. It is important to note the relationship between leaf disease and moisture. Moisture is an essential component in many leaf diseases and how they develop. A longer duration of soaking the leaves increases the risk of infection [7, 8]. Water-soaked spots on leaves are frequently early signs of plant disease. Such marks may induce cell turgor loss and tissue maceration, further developing entry areas for pathogens [7]. On

the other hand, water deficiency can also increase susceptibility to pathogens [9].

We used our in-house developed THz system to take the THz pictures of the leaves. It comprises a 0.3 THz source, a camera, and a separate digital signal processing element. With further redevelopment, the system could also become portable and be used for on-field diagnosis.

The THz camera sensor's nano bolometer thermal time constant is lower than lusec, allowing a high-frequency frame rate of the THz image at room temperature. The resolution of the THz image is dependent upon pixel size and pixel response time. Our pixel size is 1mm x 1mm. It comprises the nano bolometer, four-channel low noise amplifier (LNA) and multiplexer.

The camera is equipped with image acquisition, processing, and display software. It enables us to visualize camera-captured images in almost real-time. We have also developed several unique features, such as image enhancement, filtering, and analysis tools to extract relevant information from the raw data of the THz camera. Additional digital signal processing algorithms enhance our image quality and reduce noise.

Images were obtained by raster scanning the area with a dimension of 100 x 100 positions corresponding to the image's pixels. Data was acquired using a fast, freely on-market available universal serial bus (USB) signal acquisition card. For each pixel, 3000 samples were acquired. Then, a fast Fourier transform was performed on each pixel dataset to determine the brightness of each pixel. Due to the efficient software, this digital signal processing can run on an average laptop computer.

We also fixed a THz source above a belt conveyor to investigate the amount of moisture in the plant leaves. The nanobolometer sensor array is placed below the conveyor belt. Fig. 2 presents one of our test systems for fast analysis and comparison of different leaf water patterns.



Fig. 2. THz source above the belt conveyor and the THz sensor array below the conveyor belt.

3. Results

Figs. 3-5 show some observed leaves of the "prunus laurocerasus" or laurel plant with some disorders. In Fig. 3, we can see a healthy leave.



Fig. 3. Image of the healthy "prunus laurocerasus" or laurel plant leaf with corresponding Thz image on the right.



Fig. 4. Laurel plant leaf in a drying process with corresponding Thz image on the right.



Fig. 5. Withered laurel plant leaf with corresponding Thz image on the right.

Fig. 4 shows the leaf in the process of drying. The dryer the area of the leaf is, the lighter gray is the color in the THz image. On the other hand, darker regions are more prosperous with moisture. In Fig. 5, we can see a withered laurel plant leaf. Although there is still a green area in the left picture, the leaf is almost dehydrated, except for the leaf veins.

The advantage of THz spectroscopy for leaf internal moisture detection is the non-destructiveness of the measurement, which can preserve the sample for

further growth and analysis. Water's strong absorption of the THz waves can be detected accurately and quickly.

4. Conclusions

Different leaf diseases are an essential challenge in agriculture, landscaping, and gardening. Early detection, before visual leaf defects appear, may give us a proper diagnosis to implement management strategies necessary to maintain plant health. Maintaining optimal watering conditions through appropriate irrigation practices and environmental planning is crucial for enhancing plant resilience against diseases.

Using non-invasive THz technology and machine learning [10], it is possible to detect early signs of leaf disease on various plants and take appropriate actions to promote overall plant health. The integration of THz with DSP and imaging techniques in biology is, in our opinion, a significant advancement for the non-invasive detection of sub-surface moisture of plant leaves and for making the correct conclusions about plant conditions.

Our THz leaf monitoring experiments are currently not yet focused on research for detecting and treating specific leaf diseases.

References

[1]. L. Bin, Z. Xiao, W. Rong, M. Yu, M. Jianjun, Leaf water status monitoring by scattering effects at terahertz frequencies, *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, Vol. 245, 2021, 118932.

- [2]. N. Pengcheng, Q. Fangfang, L. Lei, D. Tao, H. Yong, S. Yongni, Z. Yi, Detection of water content in rapeseed leaves using terahertz spectroscopy, *Sensors*, Vol. 17, Issue 12, 2017, 2830.
- [3]. J. Trontelj, A. Švigelj, J. Trontelj, Novel THz sensor introduction, manufacturing description and examples of use, *Sensors & Transducers Journal*, Vol. 265, Issue 1, May 2024, pp. 1-8.
- M. Maček, J. Trontelj, A. Sešek, Bolometric Detection System with Reflecting Cavity, UK Patent GB 2513170 B, 2020-07-08, United Kingdom Intellectual Property Office, 2020.
- [5]. L. Qi, L. Minkevičius, A. Urbanowicz, A. Švigelj, I. Grigelionis, I. Kašalynas, J. Trontelj, G. Valušis, Antenna-coupled titanium microbolometers: application for precise control of radiation patterns in terahertz time-domain systems, *Sensors*, Vol. 21, Issue 10, May 2021, 3510.
- [6]. J. Trontelj, A. Švigelj, J. Trontelj ml., Novel THz sensors, in Proceedings of the 4rd Winter IFSA Conference on Automation, Robotics & Communications for Industry 4.0/5.0 (ARCI'24), 2024, pp. 43-45.
- [7]. K. Aung, Y. Jiang, S. Y. He, The role of water in plantmicrobe interactions, *The Plant Journal: for Cell and Molecular Biology*, Vol. 93-4, 2018, pp. 771-780.
- [8]. R. Li, Y. Lu, J. M. R. Peters, et al., Non-invasive measurement of leaf water content and pressure-volume curves using terahertz radiation, *Scientific Reports*, Vol. 10, 2020, 21028.
- [9]. V. Gorshkov, I. Tsers, Plant susceptible responses: the underestimated side of plant-pathogen interactions, *Biological Reviews of the Cambridge Philosophical Society*, Vol. 97-1, 2022, pp. 45-66.
- [10]. M. Koumans, D. Meulendijks, H. Middeljans, D. Peeters, J. C. Douma, D. V. Mechelen, Physics-assisted machine learning for THz spectroscopy: sensing moisture on plant leaves, *Scientific Reports*, Vol. 14, 2024, 7034.

(020)

Monitoring OoD Prediction Error in Semantic Segmentation Networks via Temporal Consistency of Logits

Youssef Shoeb ^{1,2}, Azarm Nowzad ¹ and Hanno Gottschalk ²

¹Continental AG, Germany ²Technische Universität Berlin, Germany E-mail: youssef.shoeb@continental.com

Summary: Out-of-distribution (OoD) detection is essential for robust perception in safety-critical applications such as autonomous driving. A common approach for OoD detection is to threshold the pixel-wise SoftMax entropy of a semantic segmentation model. However, such methods struggle to identify unknown objects without explicit outlier supervision. Entropy-based outlier supervision often fails to distinguish between truly unknown instances and ambiguous or low-confidence pixels that fall on the border of objects. In this work, we propose a post-processing method for reducing false-positive OoD detections in semantic segmentation networks by leveraging the temporal consistency of predictions. We observe that ambiguous pixels tend to fluctuate between semantically similar classes over time, while true OoD objects exhibit more random behavior. By filtering out transient fluctuations in predictions, our approach significantly suppresses false detections caused by ambiguity. Experiments show that incorporating temporal information reduces false-positive detections, enhancing the reliability in real-world scene understanding.

Keywords: Out-of-distribution detection, Pixel tracking, Temporal consistency, Entropy maximization, Autonomous driving.

1. Introduction

In the initial stages of developing machine learning applications, human experts collect and label large amounts of data tailored to their desired use case. This data is typically split into a training, validation and test set. The model is trained on the training set, the validation set is used to optimize the hyperparameters and prevent overfitting to the training data, and the test set is used to provide an unbiased estimate of the model's generalization performance on unseen data. The performance on the unseen data is then the expected performance during deployment. For this construction to be valid, the data distributions in the training, validation and test sets should be representative of the same underlying distribution. This requirement is often satisfied by collecting sufficient amount of data from the domain the model is expected to deploy in, and randomly splitting the data between all the sets. However, in open-set applications such as autonomous driving, where the data distribution is long-tailed, encountering out-of-distribution (OoD) data during deployment is inevitable.

State-of-the-art neural networks have been shown to produce unreliable estimates for OoD data. In safety-critical applications, it is therefore essential for machine-learning-based systems to detect OoD inputs. In the context of environmental perception, identifying and classifying objects in a scene is crucial; not only to recognize that a scene contains OoD data but also to precisely localize the regions where OoD objects appear. To address this, a rich line of research has emerged around OoD semantic segmentation [1-3], which enables pixel-level classification, allowing systems to detect and localize OoD objects within a scene. However, existing OoD detection approaches primarily rely on spatial information from individual frames, using metrics such as entropy or energy scores, without incorporating temporal features from sequential data. In applications like autonomous driving, where the input is typically a video sequence rather than a single frame, this limitation can lead to transient false positives – where momentary sensor noise or occlusions cause pixels to be misclassified as OoD. Leveraging temporal information is therefore the logical next step to enhance detection robustness, filtering out fluctuating anomalies and reinforcing consistent predictions over time.

In this work, we propose a post-processing approach to reduce false-positive predictions in OoD semantic segmentation networks by monitoring the network's temporal behavior. First, we identify OoD segments in individual frames and track their changes over time using pixel-level tracking. Then, by analyzing temporal variations in confidence scores, we differentiate between true OoD segments and false positives. We observe that true OoD segments typically exhibit rapid and random fluctuations in predictions, whereas ambiguous segments tend to fluctuate among semantically related classes (e.g., road and sidewalk). By exploiting this difference, our method significantly reduces false-positive detections, enhancing the robustness and reliability of OoD segmentation networks.

2. Method

2.1. Per-frame Out-of-distribution Detection

A neural network with a SoftMax output layer can be seen as a statistical model $f_{\theta}(y|x)$ that provides a probability distribution over n class labels $y \in C = y_1, ..., y_n$ for each pixel in the image, given the parameters θ and the input *x*. The softmax entropy,

$$E(f_{\theta}(x)) = -\sum_{y \in C} f_{\theta}(y|x) \log(f_{\theta}(y|x)) \quad (1)$$

is a widely used metric for measuring the uncertainty of a model and is a common choice for out-of-distribution (OoD) detection [4]. However, deep neural networks are often uncalibrated and produce overconfident predictions on OoD inputs. One approach to mitigate this is to use "known unknowns" as a proxy for OoD objects and explitly train the network to maximize entropy on unknown objects [5]. In this work, we use a pretrained EfficientVit model [6] trained on Cityscapes and finetune it for entropy maximization. The objective function during finetuning is defined as:

$$\mathcal{L} = (1 - \lambda) E_{\mathbb{D}_{in}(x,y)} [l_{in}(f_{\theta}(x), y)] + \\ + E_{\mathbb{D}_{out}(x)} [l_{out}(f_{\theta}(x))], \lambda \in [0,1],$$

$$(2)$$

where l_{in} is the negative log-likelihood for the in-distribution target class, and l_{out} is the negative log-likelihood averaged across all classes for proxy OoD samples pasted onto in-distribution images.

During inference, the SoftMax entropy is computed for each pixel in the image. A pixel is classified as an OoD candidate if its entropy exceeds a predefined threshold. To identify coherent OoD regions, we define a segment as a connected component of OoD pixels that share a contiguous boundary with in-distribution pixels. To facilitate the temporal consistency and analysis, the centroid of each segment is tracked across frames. The probability distribution of a segment is then the average of all the individual pixels in a segment.

2.2. Point Tracking

Estimating motion in image sequences was traditionally performed using optical flow [7] or feature tracking [8]. Optical flow estimates a dense motion between a pair of frames, and feature tracking follows a sparse set of points over multiple frames.

The particle video algorithm [9] is a middle-ground between both approaches and produces motion estimates that are both spatially dense and temporally long-range. Several neural network components have then been used to extend the particle video algorithm to track the trajectories of any individual pixels independently [10], in a task referred to as *tracking any point*. This advancement enables fine-grained motion estimation, allowing for robust tracking in complex, dynamic environments.

Formally, given a sequence of N frames $[f_t]_{t=0}^N \in \mathbb{R}^{3 \times H \times W}$, and a query point $q_0 = (x_0, y_0)$ where (x_0, y_0) is the initial location of the query point, the goal is to predict the corresponding point track $p_t = (x_t, y_t) \in \mathbb{R}^2$ fort = 1, ..., N. To obtain the estimate p_t , we compute convolutional features for

every frame in the sequence f, and then sample the correlations between the features around the query frame and all other points. For further details on the point tracking method used in this work, we refer the reader to [11].

In this work, we extend point tracking to follow the motion of detected OoD segments. Specifically, the centroid of each OoD segment is initialized as a query point, and its trajectory is estimated across the subsequent frames. This allows for the tracking of moving OoD objects in video sequences even through occlusions.

2.3. Temporal Out-of-distribution Detection

To evaluate the temporal consistency of the model's predictions across consecutive frames, we use the *mean temporal consistency* (mTC) metric This metric measures the average distance between predictions for consecutive frames. For a sequence of T frames, mTC is defined as:

$$mTC = \frac{1}{T-1} \sum_{t=1}^{T-1} d_t, \qquad (3)$$

where d_t represents the distance between the model's predictions for two consecutive frames x_t and x_{t+1} . This distance is calculated using the Wasserstein distance:

$$d_{t} = \int_{-\infty}^{\infty} w(z) |P_{t}(z) - P_{t+1}(z)| \, dz, \qquad (4)$$

where $P_t(z)$ and $P_{t+1}(z)$ denote the predicted probability distributions for frames t and t+1, respectively. The class-specific weight function w(z)reflects the semantic similarity between classes, assigning lower weights to transitions between semantically similar classes and higher weights to transitions between semantically distant classes.

In this work, we use the hierarchical structure of the label ontology to define the weight function w(z). Specifically, we assign a weight of zero to transitions occurring within the same superclass (*e.g.*, road and sidewalk), this ensures that minor intra-category variations that occurs due to expected uncertainty do not contribute to the measured inconsistency. Conversely, transitions across different superclasses are assigned an equal weight of one, emphasizing significant semantic shifts. This hierarchical weighting strategy ensures that the mTC metric robustly captures meaningful temporal inconsistencies while remaining invariant to fine-grained within-class variations.

3. Experiments

To evaluate the performance of our proposed method we require video sequences with consecutive frames for tracking. We Evaluate our method on the Street Obstacle Sequences (SOS) dataset [12]. The SOS dataset contains 20 video sequences of street scenes with more than a thousand labelled frames that include up to two OoD objects.

Table 1 shows the result of our method compared to only using Entropy Maximization for OoD detection. Our method improves overall performance, increasing the F1 score from 58.78 to 62.92. This improvement is primarily driven by a substantial reduction in false positives (FP), which decreased by approximately 50 % from 1135 to 564. However, it's important to note that this improvement comes at the cost of a slight decrease in true positives (TP). Our method detects 940 true positives compared to 1135 by the Entropy Maximization approach, representing a decrease of about 17 %. This occurs because some of the OoD objects don't exhibit a high *mTC* score as shown Fig. 1.

 Table 1. Performance on SOS Dataset.



Fig. 1. Comparison between the range of *mTC* values for OoD objects and false positives.

4. Conclusion

In this work, we proposed a method for reducing false positive predictions in OoD semantic segmentation networks by analyzing the network's temporal behavior. By integrating temporal information and tracking OoD segments over time, our approach enhances the overall performance of OoD segmentation. Specifically, we explored mean temporal consistency (mTC) as a metric for filtering false positives, demonstrating a significant reduction in false positives. Future work could investigate additional temporal features, such as object motion patterns, to further refine detections. Adaptive thresholding based on scene dynamics or segment properties, as well as hybrid approaches that combine multiple post-processing strategies, are also promising research directions for improving robustness and generalization.

Acknowledgements

The research leading to these results is funded by the German Federal Ministry for Economic Affairs and Climate Action within the project "just better DATA".

References

- R. Chan, M. Rottmann, H. Gottschalk, Entropy maximization and meta classification for out-of-distribution detection in semantic segmentation, in *Proceedings of the IEEE/CVF Int. Conference Comput. Vis. (ICCV'21)*, 2021, pp. 5128-5137.
- [2]. N. Nayal, Y. Shoeb, F. Güney, A likelihood ratio-based approach to segmenting unknown objects, *arXiv* preprint, 2024, arXiv:2409 06424.
- [3]. Y. Shoeb, N. Nayal, A. Nowzad, F. Güney, H. Gottschalk, Segment-level road obstacle detection using visual foundation model priors and likelihood ratios, *arXiv preprint*, 2024, arXiv: 2412.05707.
- [4]. D. Hendrycks, K. Gimpel, A baseline for detecting misclassified and out-of-distribution examples in neural networks, in *Proceedings of the Int. Conference Learn. Represent.*, July 2022.
- [5]. Y. Shoeb, A. Nowzad, H. Gottschalk, Out-of-distribution segmentation in autonomous driving: problems and state of the art, *arXiv preprint*, 2025, arXiv:2503.08695.
- [6]. H. Cai, J. Li, M. Hu, C. Gan, S. Han, EfficientViT: Multi-scale linear attention for high-resolution dense prediction, arXiv preprint, 2022, arXiv:2205.14756.
- [7]. S. S. Beauchemin, J. L. Barron, The computation of optical flow, ACM Comput. Surv., Vol. 27, Issue 3, 1995, pp. 433-466.
- [8]. J. Shi, Good features to track, in *Proceedings of the IEEE Conference Comput. Vis. Pattern Recognit.* (CVPR'94), June 1994, pp. 593-600.
- [9]. P. Sand, S. Teller, Particle video: Long-range motion estimation using point trajectories, *Int. J. Comput. Vis.*, Vol. 80, 2008, pp. 72-91.
- [10]. A. W. Harley, Z. Fang, K. Fragkiadaki, Particle video revisited: Tracking through occlusions using point trajectories, in *Proceedings of the Eur. Conference Comput. Vis. (ECCV'22)*, Oct. 2022, pp. 59-75.
- [11]. N. Karaev, I. Rocco, B. Graham, N. Neverova, A. Vedaldi, C. Rupprecht, CoTracker: It is better to track together, in *Proceedings of the Eur. Conference Comput. Vis. (ECCV'24)*, 2024, pp. 18-35.
- [12]. K. Maag, R. Chan, S. Uhlemeyer, K. Kowol, H. Gottschalk, Two video data sets for tracking and retrieval of out-of-distribution objects, in *Proceedings* of the Asian Conference Comput. Vis. (ACCV'22), 2022, pp. 3776-3794.

(021)

Examining Physiological Responses to Misophonic Triggers

C. O'Reilly ¹⁻⁴, X. Yang ^{2,5}, S. Oh ^{2,5}, D. Wedell ^{2,5} and <u>S. V. Shinkareva</u> ^{2,5}

¹ Department of Computer Science and Engineering, University of South Carolina, Columbia, SC, USA
 ² Institute for Mind and Brain, University of South Carolina, Columbia, SC, USA
 ³ Artificial Intelligence Institute, University of South Carolina, Columbia, SC, USA
 ⁴ Carolina Autism and Neurodevelopment Research Center, University of South Carolina, Columbia, SC, USA
 ⁵ Department of Psychology, University of South Carolina, Columbia, SC, USA
 E-mail: christian.oreilly@sc.edu; shinkareva@sc.edu

Summary: We collected and analyzed face electromyogram, skin electrodermal activity, peripheral skin temperature, and electrocardiogram in 60 participants with (N = 35) and without (N = 25) misophonia, a condition characterized by decreased tolerance to innocuous sounds. Our goal was to characterize the physiological response to misophonia-triggering sounds objectively. We found that misophonic responses can be identified in some cases through atypical physiological reactions to triggering stimuli, though not all participants exhibited this response. Our analyses suggest a large interindividual variability in response to misophonic triggers and highlight the need for methodological adjustments in future experiments to increase the detectability of misophonic reactions.

Keywords: Biosignals, Misophonia, Electromyogram, Electrodermal activity, Electrocardiogram, Peripheral skin temperature, Machine learning.

1. Introduction

Misophonia is a condition characterized by a decreased tolerance to innocuous sounds (i.e., triggers), such as chewing, which can be highly debilitating for some individuals [1]. Nearly 5 % of the adult population is affected by misophonia [2, 3]. Despite its prevalence, research on misophonia remains limited, and the biological mechanisms underlying this condition are poorly understood. This gap in knowledge hinders the development of effective treatments and preventive interventions. Current approaches to studying misophonia often rely on participants' self-report ratings of distress in response to triggers, which are subjective and prone to variability within participants. To address this issue, identifying misophonic triggers through physiological signals offers a more objective means of confirming responses and assessing the severity of this condition. Electrodermal activity (EDA) and heart rate have previously been used to describe misophonic responses [4-7]. This study aims to further characterize the physiological responses associated with misophonia by also including facial electromyography (EMG) and peripheral skin temperature (SKT), two recording modalities that have not been previously explored in this condition.

2. Methods

Experimental paradigm. We recorded five physiological signals: EMG for the corrugator supercilii (EMG C) and zygomaticus major (EMG Z) muscles, EDA, SKT (measured on the palm side of the thumb of the left hand), and electrocardiogram (ECG), at 1,000 Hz¹ [8] in participants with and without misophonia. Participants listened to sounds, viewed silent videos associated with sounds, or were asked to think about sounds in a 3 (sensory modality: auditory, visual, mental imagery) by 3 (stimuli: trigger, aversive, non-aversive) factorial design. Participants rated how distressing (misophonia group) or antisocial (control group) the stimulus was for each trial (Fig. 1). This distinction between groups is necessary because distress is a key feature of misophonic responses, which is not typically experienced by control participants, even though they may find the stimuli unpleasant or antisocial. Participants aged 18 to 45 without a history of neurological or psychiatric diagnoses, hearing loss, or hyperacusis were drawn from the South Carolina community and compensated for their time. The Misophonia group (N = 35, 30 female; $M_{age} = 25.3$, $SD_{age} = 8.10$) consisted of individuals who experience misophonic reactions triggered by sounds with distinct visual components but not associated with specific individuals. Trigger sets for each misophonia participant were selected

¹ Before preprocessing, six recordings acquired at 200 Hz and two at 2,000 Hz were resampled at 1,000 Hz using the MNE-Python resample() function, which adopts the same

approach as SciPy, relying on the Fast Fourier Transform. We observed no systematic effects of this resampling on classification.

based on the categories of triggers identified during the screening interviews. Control group participants $(N = 25, 20 \text{ female}; M_{age} = 25.0, SD_{age} = 8.10; \text{ the}$ recruitment of control participants is ongoing) were matched with misophonia participants on age, biological sex, handedness, and stimuli presentation. Control participants were also screened to ensure they did not have misophonia. All participants completed misophonia severity questionnaires, two the Duke-Vanderbilt Misophonia Screening Questionnaire (DVMSQ) [9] and the Selective Sound Sensitivity Syndrome Scale (S-Five) [3]. They also filled two mental imagery questionnaires, the Vividness of Visual Imagery Questionnaire (VVIQ) [10] and the Bucknell Auditory Imagery Scale (BAIS) [11], which included two subscales for vividness (BAIS-V) and control (BAIS-C). The study was carried out in accordance with the procedures and protocols approved by the University of South Carolina Institutional Review Board, and all participants signed an informed consent.

Two groups

misophonia, control: matched on sex, age, handedness & stimuli

Nine experimental conditions



Fig. 1. Experimental design.

Preprocessing and feature computation. EMG signals were notch-filtered at 60 Hz and band-pass filtered in the 20-500 Hz range. They were then rectified, baseline-corrected using the average amplitude over one second before the stimulus onset, and smoothed with a rolling average (window size = 100 samples, or 0.1 s). From these preprocessed signals, we extracted the peak amplitude within the 7-second window following the stimulus as features for classification. EDA and temperature were similarly baseline-corrected and smoothed. For these signals, the amplitude 15 seconds after stimulus onset was used as a feature. For the ECG signal, R peaks were automatically detected using AcqKnowledge (Biopac Systems, Inc., Goleta, CA, USA). Two research assistants manually edited the R peaks to correct motion artifacts and misclassifications. The instantaneous heart rate was computed as the inverse of the R-R interval and used to characterize the ECG signals. We used the mean-square-root of the baselinecorrected heart rate within a 2-7 second window following stimulus onset as a feature. The defined windows were based on preliminary analyses. EMG Z

data from one control participant were excluded due to issues related to the data acquisition process.

Classical statistical analysis. Student's t-tests and correlation analysis were used to compare stimulus types within individuals for each physiological signal. The correlation between ratings and physiological responses was assessed using Pearson's correlation across all trials, regardless of stimulus type or sensory modality. All tests were conducted at an alpha level of 0.001 without correction for multiple comparisons. For statistical significance between time series, multiple comparisons were corrected using a cluster-level statistical permutation test, as implemented in MNE-Python [12]. Since not all trigger trials led to misophonic responses, we selected only half of the trigger trials based on the higher distress/antisocial ratings reported by the participants. This selection ensures that the average physiological response to triggers is not diluted by trials in which participants did not subjectively feel triggered. Since preliminary analyses showed a weaker response for the mental imagery, it has been excluded from the classical statistical analyses. We report on the relative strength of responses by modality in the machine learning analysis.

Machine learning analysis. We used machine learning to determine whether physiological responses could predict group membership (i.e., misophonia versus control) and stimulus type (i.e., trigger, aversive, and non-aversive). Linear Support Vector Machines with default parameterization were employed (using the LinearSVC implementation in Scikit-Learn [13], based on LIBLINEAR [14]). A leave-one-out cross-validation was used to assess group classification, and a leave-one-group-out cross-validation, with the participant serving as the grouping factor, was used for assessing stimulus type classification. Features were standardized by removing the mean and scaling to unit variance before classification. For stimulus classification, both accuracy and weighted f1 scores were used to measure performance for the balanced case. For the unbalanced case of group classification, only weighted f1 scores were reported. To assess the statistical significance of fl scores and accuracies, we bootstrapped the trial selection within participants 100 times using random selection with full sample size and replacement. We evaluated the relative importance of the different features in the classification using permutation feature importance [15]. We imputed missing data (i.e., features from the excluded EMG Z channel for one control participant) using the nearest neighbors imputation approach [16] provided by the scikit-learn KNNImputer class.

3. Results

First, we examined individual differences in physiological responses to trigger versus aversive stimuli for each signal. We found that misophonia triggers could be identified in some participants (Fig. 2A) and at the group level (Fig. 2D) based on physiological responses. Participants who exhibited significant physiological responses to triggers were generally those who self-reported experiencing the most distress when presented with these triggers. This is demonstrated by the fact that all but one significant physiological response differences at the individual level (Fig. 2A) were observed in the nine participants who showed a clear difference in ratings between aversive and trigger stimuli, with t-statistics greater than 10 (Fig. 2C). Additionally, self-report and physiological responses were strongly correlated in a large proportion of participants, particularly within the misophonia group (Fig. 2B).

To further investigate the predictive value of physiological responses, we used machine learning to

assess whether group membership and stimulus types could be predicted from physiological data. When all features of the physiological responses were considered together, participant classification into misophonia and control groups was not statistically significant. Not surprisingly, self-report distress/antisocial ratings were strong predictors of group membership (Fig. 3A). However, predictive information was found in physiological responses when contrasts between stimulus types were used (Fig. 3B). The contrast between aversive and non-aversive stimuli appeared to be as indicative of misophonia as contrasts involving trigger stimuli. Additionally, mental imagery contrasts between stimulus types seemed to provide as much predictive information as perceived auditory or visual stimuli.



Fig. 2. Detecting triggers from physiology. A) Significant *t* values from Student t-tests comparing trigger and aversive responses by physiological signal and participant. B) Significant Pearson's correlations between physiological responses and distress (misophonia) or antisocial (control) ratings. C) Column 1: Student's *t* values for the difference in means between distress/antisocial ratings for triggers (T) and aversive (Av) stimuli, noted t(T, Av). Columns 2-4: Average distress/antisocial ratings for T, Av, and NAv (non-aversive) stimuli. Columns 5-6: DVMSQ and S-Five misophonia severity scores. For A-C, the dashed light red line separates misophonia and control participants. Within groups, participants are sorted by decreasing t(T, Av) values. The dashed light gray line identifies misophonia participants with t(T, Av) > 10. D) Physiological response to stimuli. Shaded regions represent the 95 % bootstrapped confidence intervals. Black overlines show time intervals where differences between responses to trigger and aversive stimuli were statistically significant according to a cluster-based permutation test.

On average, physiological responses to auditory and visual but not mental imagery trials were predictive of stimulus types (Fig. 3C, E). However, these classification results only held for a subset of participants with misophonia (Fig. 3F), similar to results in Fig. 2A. These results were primarily driven

by responses to triggers in misophonia participants (Fig. 3G, H), particularly those who self-reported being strongly triggered (Fig. 3I). Interestingly, there was greater confusion between triggers and non-aversive stimuli than between triggers and aversive stimuli, suggesting that in some cases, misophonic responses might not have been triggered as expected. Alternatively, if ECG was driving this pattern in the confusion matrices (Fig. 3G-I), the ECG response for non-aversive stimuli being midway between the response for triggers and aversive stimuli (Fig. 2D) could also explain this observation. To test that possibility, we evaluated the relative importance of the classification features. In line with the results displayed in Fig. 2, EMG C was the most important

feature, followed by SKT and ECG (Fig. 4). Further, we performed an ablation study where we removed the features derived from ECG and the confusion pattern did not change. Thus, it appears unlikely the difference in ECG response to the different stimulus types is responsible for this observation.

Finally, we assessed the relationship between prediction accuracies for classifying stimulus types (Fig. 3D, F) and misophonia severity scores (Fig. 2C) across participants. As expected, we found a significant correlation between these measures (Fig. 5). However, contrary to our expectations, we did not observe a significant correlation between prediction accuracies and mental imagery scores in the mental imagery condition.



Fig. 3. Predictivity of the physiological response assessed through machine learning. A) Prediction (weighted f1 scores) of group membership (misophonia vs. controls) for the average response per sensory modality and stimulus type using physiology data only (i.e., 60 participants; 45 features: 3 stimulus types \times 3 sensory modalities \times 5 physiological signals), self-report ratings only (9 features), or both (54 features). B) Prediction (weighted f1 scores) of group membership for each sensory modality (V: visual; A: auditory; I: mental imagery) and stimulus type contrasts (T: trigger; Av: aversive; NAv: non-aversive). C, E) Weighted f1 scores (C) and accuracies (E) for stimulus type prediction per sensory modality. D, F) Prediction (weighted f1 scores) for the stimulus types per participant in the control group (D) and the misophonia group (F). Participants were sorted by decreasing order of accuracy. Participant pairing between groups is indicated using the same numbers. For panels A-F, the pale dashed lines indicate the chance level. Participant numbers were prefixed with V, A, or I when the 5th percentile of the bootstrapped accuracy distribution was above chance levels (0.33). G-I) Confusion matrices per sensory modality for predicting stimulus types in the control (G), the misophonia (H) groups, and the misophonia participants with *t*(T, Av) > 10 as defined in Fig. 2 (I), scaled so that the chance level is 1.0.

4. Discussion

Triggers were robustly identifiable from physiological responses for a relatively small subset of participants with misophonia (Fig. 2A). The absence of significant physiological responses to triggers in some misophonia sufferers may be due to factors such as the failure to replicate the triggering context (e.g., chewing sound in the library) [7], the idiosyncratic nature of the stimuli (e.g., the wrong type of chewing), or insufficient trigger duration. Participants with the highest severity scores were not always the ones for whom triggers were identified, suggesting that we may not have elicited the expected response in all participants. Previous studies used longer stimuli (e.g., 15 s [5]), and as shown in Fig. 2D, physiological responses were still building at the end of the 5-second stimulus.

EMG C -	3.62	2.30	1.37
EMG Z -	0.06	-0.00	-0.25
EDA -	-0.10	0.01	-0.34
SKT -	0.45	0.48	-0.13
ECG -	0.78	0.68	0.03
	Å	v	i

Fig. 4. Feature importance for stimulus type prediction by modality. Larger numbers indicate features with a higher influence on the classification.

		All		Control		Mis	opho	nia	
DVMSQ -	0.08	0.02	0.81	-			0.08	0.02	0.81
S-Five -	0.03	0.03	0.41	- 0.26	0.76	0.57	- 0.93	0.40	0.51
BAIS-C -	0.68	0.70		- 0.61	0.38	0.98	0.19		0.10
BAIS-V -	0.97	0.59	0.44	- 0.30	0.64	0.65	- 0.23		0.15
VVIQ -	0.46	0.92	0.91	- 0.43	0.75	0.46	- 0.61	0.90	0.86
	Å	v	i	Å	v	i	Å	v	i

Fig. 5. P-values of Pearson's correlations between stimulus type prediction accuracy (Fig. 3D, F) and both misophonia (Fig. 2C) and mental imagery scores.

Self-report ratings were sufficient to identify group membership, though not perfectly (Fig. 3A). This could be attributed to several factors: misophonia participants may not have been triggered by the designated trigger stimuli, self-report measures may have been noisy or unreliable, or the classifier may not have fully captured the complexity of the data. The differences in physiological responses between the stimuli in the misophonia group were more pronounced and driven primarily by trigger stimuli (Fig. 3C-I). Additionally, physiological responses to both aversive and non-aversive non-trigger stimuli differed between individuals with misophonia and those without (Fig. 3B), suggesting that misophonia is associated with a broader intolerance to sound [17-19]. In future work, we plan to explore using autoencoders to extract richer features and apply pre-trained deeplearning models to the time-frequency representation of these physiological signals to improve our ability to identify individual misophonic triggers.

5. Conclusion

Physiological data from multiple recording modalities offer valuable insights into misophonia, highlighting distinct response patterns to trigger and non-trigger stimuli. These physiological responses differ between individuals with misophonia and those without, indicating a clear distinction in how each group reacts to these stimuli. These differences extend beyond sounds to include silent videos and even the mental imagery of sounds. Together, these findings suggest that misophonia may be identifiable through unique physiological patterns, regardless of the stimulus type, highlighting the potential of physiological measures for improving our understanding and diagnosis of misophonia.

Acknowledgment

This work was supported by a grant from the Misophonia Research Fund.

References

- S. E. Swedo, et al., Consensus Definition of Misophonia: A Delphi Study, Front. Neurosci., Vol. 16, 2022, https://www.frontiersin.org/journals/ neuroscience/articles/10.3389/fnins.2022.841816
- [2]. L. J. Dixon, M. J. Schadegg, H. L. Clark, C. J. Sevier, S. M. Witcraft, Prevalence, phenomenology, and impact of misophonia in a nationally representative sample of U.S. adults, *J. Psychopathol. Clin. Sci.*, Vol. 133, Issue 5, Jul. 2024, pp. 403-412.
- [3]. S. Vitoratou, C. Hayes, N. Uglik-Marucha, O. Pearson, T. Graham, J. Gregory, Misophonia in the UK: Prevalence and norms from the S-Five in a UK representative sample, *PLoS ONE*, Vol. 18, Issue 3, Mar. 2023, e0282777.
- [4]. A. Schröder, et al., Misophonia is associated with altered brain activity in the auditory cortex and salience network, *Sci. Rep.*, Vol. 9, Issue 1, May 2019, 7542.
- [5]. S. Kumar, *et al.*, The brain basis for misophonia, *Curr. Biol.*, Vol. 27, Issue 4, Feb. 2017, pp. 527-533.
- [6]. M. Edelstein, D. Brang, R. Rouw, V. S. Ramachandran, Misophonia: physiological investigations and case descriptions, *Front. Hum. Neurosci.*, Vol. 7, Jun. 2013, 296.
- [7]. M. Siepsiak, S. R. Vrana, A. Rynkiewicz, M. Z. Rosenthal, W. Ł. Dragan, Does context matter in misophonia? A multi-method experimental investigation, *Front. Neurosci.*, Vol. 16, Jan. 2023.
- [8]. S. Oh, X. Yang, W. M. Hayes, A. Anderson, D. H. Wedell, S. V. Shinkareva, Physiological responses to aversive and non-aversive audiovisual, auditory, and visual stimuli, *Biol. Psychol.*, Vol. 195, Jan. 2025, 108994.
- [9]. Z. J. Williams, C. J. Cascio, T. G. Woynaroski, Psychometric validation of a brief self-report measure of misophonia symptoms and functional impairment: The Duke-Vanderbilt misophonia screening questionnaire, *Front. Psychol.*, Vol. 13, 2022, 897901.
- [10]. D. F. Marks, New directions for mental imagery research, J. Ment. Imag., Vol. 19, Issue 3-4, 1995, pp. 153-167.
- [11]. A. R. Halpern, Differences in auditory imagery selfreport predict neural and behavioral outcomes, *Psychomusicology Music Mind Brain*, Vol. 25, Issue 1, 2015, pp. 37-47.
- [12]. A. Gramfort, et al., MEG and EEG data analysis with MNE-Python, Front. Neurosci., Vol. 7, 2013.
- [13]. F. Pedregosa, et al., Scikit-learn: Machine learning in Python, J. Mach. Learn. Res., Vol. 12, Issue 85, 2011, pp. 2825-2830.

- [14]. R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, C.-J. Lin, LIBLINEAR: a library for large linear classification, *J. Mach. Learn. Res.*, Vol. 9, Jun. 2008, pp. 1871-1874.
- [15]. L. Breiman, Random Forests, *Mach. Learn.*, Vol. 45, Issue 1, Oct. 2001, pp. 5-32.
- [16]. O. Troyanskaya, et al., Missing value estimation methods for DNA microarrays, *Bioinformatics*, Vol. 17, Issue 6, Jun. 2001, pp. 520-525.
- [17]. N. Andermane, M. Bauer, E. Sohoglu, J. Simner, J. Ward, A phenomenological cartography of

misophonia and other forms of sound intolerance, *iScience*, Vol. 26, Issue 4, Apr. 2023, 106299.

- [18]. N. Andermane, M. Bauer, J. Simner, J. Ward, A symptom network model of misophonia: From heightened sensory sensitivity to clinical comorbidity, *J. Clin. Psychol.*, Vol. 79, Issue 10, 2023, pp. 2364-2387.
- [19]. H. A. Hansen, A. B. Leber, Z. M. Saygin, What sound sources trigger misophonia? Not just chewing and breathing, *J. Clin. Psychol.*, Vol. 77, Issue 11, 2021, pp. 2609-2625.

(022)

Comparative Study of Route Algorithms Applied to Drones

Jezabel Molina-Gil¹, Ricardo Aguasca-Colomo² and José Gregorio Dorta-Luis¹

 ¹ University of la Laguna (ULL). Department of Computer Science and Systems, San Cristobal de la Laguna, 38200, Tenerife, Spain
 ² Instituto Universitario de Sistemas Inteligentes y Aplicaciones Numéricas en Ingeniería, University of Las Palmas de Gran Canaria, 35017 Gran Canaria, Spain E-mail: jmmolina@ull.edu.es

Summary: This research explores the implementation of advanced path-planning algorithms, enhanced by Artificial Intelligence (AI), to optimize drone missions for rescue and exploration purposes. By leveraging AI-driven heuristics, the study addresses key challenges such as route efficiency, environmental adaptability, and resource utilization. The developed application, integrates user-friendly interfaces with interactive mapping, enabling customization of mission parameters like wind conditions, altitude, and drone autonomy. Tested algorithms, including Nearest Neighbor Traversal and its variations, demonstrated diverse performance strengths, with AI playing a pivotal role in overcoming local optima and improving overall results. The research highlights the transformative potential of combining AI with drone technology to enhance operational efficiency in critical scenarios. Future work aims to address computational limitations and expand adaptability for complex environments.

Keywords: Route planning, Drones, Routing algorithms, Wind, Drone autonomy, TSP, Rescue, Exploration.

1. Introduction

In recent decades, the use of drones has grown exponentially across diverse applications. These unmanned aerial vehicles have become increasingly accessible, finding use in sectors ranging from photography and filmmaking to agriculture and surveillance. The proliferation of drones is largely due to technological advancements that have reduced costs and enhanced ease of use, allowing both professionals and hobbyists to operate them for various purposes.

Modern drones are equipped with high-resolution cameras, advanced sensors, and precise navigation systems, enabling them to autonomously perform tasks with high efficiency. complex These technological advancements allow operators to define intricate flight paths using relatively simple input data, thereby eliminating the need for continuous manual intervention. The advantages of employing unmanned aerial systems are evident: by removing the constraints associated with human involvement, factors such as fatigue and operational limitations are mitigated, enabling sustained and highly efficient operations. Furthermore, this autonomy allows human operators to focus on higher-level tasks while a fleet of drones simultaneously executes multiple routes in parallel. As a result, the development of optimized route-planning strategies has become increasingly critical to maximizing the effectiveness of drone operations.

Although autonomous drone route planning offers benefits across a wide range of applications, this research focuses on the specific domain of route planning for exploration and rescue missions. The ability to define optimal routes for such operations can significantly enhance area coverage, resource allocation, and response times. This study evaluates various coverage path algorithms designed to ensure efficient coverage of the target area while minimizing the distance traveled, thereby addressing the unique challenges posed by exploration and rescue scenarios.

The primary objective of this research is to develop a software solution capable of generating optimized drone routes for rescue and exploration missions. The application is designed to provide a user-friendly interface that enables users to define a specific area of interest through an interactive map. Users can customize key parameters such as the drone's launch point, autonomy, camera field of view (FOV), altitude, speed, and wind direction.

Using these inputs, the program employs advanced coverage path algorithms and heuristic methods to generate an optimal flight path. The output is a mission file containing a detailed sequence of waypoints for the drone to follow. This file is fully compatible with widely used mission planning tools, such as Mission Planner [1] and QGroundControl [2]. These platforms act as intermediaries, providing a graphical interface for configuring and visualizing flight parameters, thereby facilitating efficient mission execution.

2. State of the Art

Path-planning algorithms have been a fundamental topic in graph theory [3], [4] and computer science for decades. Classic algorithms like Dijkstra's and Floyd-Warshall's methods [5] are well-suited for finding the shortest path between nodes in weighted graphs. However, due to their computational complexity, these algorithms are not feasible for large-scale problems. Heuristic approaches, such as the A* algorithm [6], provide efficient solutions by combining optimal pathfinding with heuristic guidance. These approaches are increasingly

augmented with AI techniques, allowing for more intelligent and adaptive solutions in dynamic environments.

For exploration and rescue missions, the goal is not only to find the shortest path between nodes but to determine the optimal route that visits all nodes exactly once without repetition. This challenge resembles the Traveling Salesperson Problem [7] (TSP), an NP-complete problem. In this research, variants of the TSP are explored to evaluate their performance in drone mission planning.

3. Problem Definition

The main problem addressed in this research is finding an optimal route that enables a drone to fully explore a specific area and generate the corresponding waypoint files. Key considerations include:

- **Defining the maximum exploration area:** Allowing users to delineate the area achievable by the drone through a graphical interface;
- Environmental factors: Integrating wind direction and speed to optimize energy consumption;
- **Route optimization:** Employing advanced algorithms to generate efficient waypoint sequences;
- File compatibility: Ensuring the output files are compatible with Mission Planner and QGroundControl.

4. Development and Implementation

4.1. Technical Factors in Route Planning

The planning process considers multiple technical factors essential for mission optimization. Among these, altitude and speed play a pivotal role. The drone's altitude directly impacts terrain coverage and data quality. A higher altitude increases the field of view (FOV), allowing larger areas to be surveyed per capture, while lower altitudes enhance the resolution of collected data. In equation (1) *d* represents the half of the base of the triangle that forms the drone's field of view at a height *h*, where FOV is the angle α .

$$d = h \cdot \cot(\alpha/2) \tag{1}$$

Speed affects mission duration and the rate at which the drone gathers information, and requires careful adjustment based on mission objectives. equation (2) represents the drone's area coverage rate, defined as the rate at which the drone can systematically cover a defined area while maintaining a constant field of view at a given altitude and flight speed. Note that for simplification, it is assumed that the FOV angle of the camera is equal in both axes, meaning the field of view is symmetric in width and length.

$$V_B = 4 \cdot d^2 / t_{BSfov} \tag{2}$$

Additionally, drone autonomy and FOV are critical. The software also calculates the maximum scanning surface for a given range and a given height of the drone. This surface is given by equation (3):

$$SB_{max} = 2 \cdot h \cdot v \cdot t_{mis} \cdot \tan(\alpha/2) \tag{3}$$

This maximum coverage area is designed for routes that follow a ladder pattern.

The planning system ensures flexibility by allowing user customization of these parameters, avoiding restrictions to specific drone models. Autonomy is treated conservatively, utilizing only 70 % of the total battery capacity to ensure safe return to the launch point. Wind direction and speed are also integrated into the planning process to enhance energy efficiency and minimize flight times Fig. 1.



Fig. 1. User Interface.

4.2. Algorithmic and Implementation Strategies

The system addresses irregular mission areas by constructing a bounding parallelogram around the user-defined exploration zone, subdividing it into smaller squares. These squares act as waypoints, with irrelevant waypoints outside the defined area eliminated through custom filtering algorithms. The key strategies include, *Grid Rotation* and *Home Point*.

- Grid Rotation, aligns the waypoint grid 45° with wind direction to reduce resistance and optimize energy use;
- Home Point starts the mission at the waypoint closest to the wind's origin direction, minimizing energy consumption during the initial traversal.

The implemented algorithms include coverage path algorithms, which calculate efficient coverage paths within the defined exploration area. Additionally, traversal heuristics, based on variants of the Nearest Neighbor algorithm, optimize the sequence of visited waypoints. These heuristics incorporate directional and randomization elements to address local optima and mitigate abrupt directional changes, further enhancing performance. This information is illustrated in Fig. 2.



Fig. 2. Three missions Grid Rotation and Home Point.

5. Results Analysis

This study evaluates three drone traversal strategies: nearest neighbor (NN), nearest directional neighbor (NDN), and nearest random neighbor (NRN). Four metrics were analyzed: total explored area, number of missions, number of waypoints, and total distance traveled, along with the impact of wind direction.

The results (Table 1) indicate that the 'Nearest Directional Neighbor Traversal' strategy minimizes traveled distance, optimizing efficiency. In contrast, the 'Nearest Random Neighbor Traversal' strategy was less efficient in this evaluation, although its variability suggests potential improvements with multiple executions.

It was found that the location of the starting point, relative to the wind direction significantly affect the total travel distance, but wind direction does not substantially affect the performance of the traversal strategies. Therefore, algorithm selection should prioritize efficiency over wind direction.

6. Conclusions

The findings indicate that the 'Nearest Directional Neighbor Traversal' strategy is the most efficient on average, as it consistently achieves the shortest travel distances. However, the 'Nearest Random Neighbor Traversal' strategy holds potential for discovering solutions through multiple executions, better leveraging its stochastic nature to escape local optima. Additionally, the starting point's location relative to the wind direction significantly impacts the total travel distance, with favorable positioning reducing distance considerably. In contrast, no clear correlation was observed between wind direction and the performance of the different traversal strategies. Thus, while wind direction itself is not a decisive factor, the strategic selection of the starting point remains crucial for optimizing drone traversal efficiency.

 Table 1. Result for large and irregular enclosure.

Strateg y	wind	m ²	Way points	Distance
NN				40737
NDN	NO	1241120	541	40578
NRN				40319
NN				38521
NDN	Ν	1218309	531	37846
NRN				38731
NN				34223
NDN	E	1218333	531	33886
NRN				34091

7. Limitations and Future Work

Despite its robustness, the program has limitations, such as computational inefficiencies for missions with more than 2000 waypoints and a simplified two-dimensional approach that does not account for altitude variations. However, the waypoint limit is not a significant constraint in most practical applications, as typical drone missions usually involve fewer waypoints. Future developments may include matrix-based waypoint management and optimization of trajectories to and from the home point, improving execution time and adaptability to complex terrains.

Additionally, future work will include benchmarking against AI-based routing models, such as Genetic Algorithms and Reinforcement Learning, to assess their potential advantages over the heuristic methods evaluated in this study.

Another limitation of this study is the lack of a detailed computational efficiency analysis for real-time execution. Future work will focus on performance benchmarking of the proposed heuristics while also validating the algorithms through real-world flight tests to assess their practical applicability and robustness under real environmental conditions.

Acknowledgements

This research is possible thanks to the following organizations and projects: SCITALA C064/23 and the Cybersecurity Chair ULL-INCIBE funded though NextGeneration EU through Recovery, Transformation and Resilience Plan, the Spanish Ministry of Universities through Project FIDEMOV, the PID2022-138933OB-I00 project, and the 2023DIG28 project supported by CajaCanarias and Fundación la Caixa.

References

- [1]. Mission Planner Home Mission Planner documentation, https://ardupilot.org/planner/
- [2]. QGC, QGroundControl Drone Control, https://qgroundcontrol.com/

- [3]. A. Gasparetto, P. Boscariol, A. Lanzutti, R. Vidoni, Path planning and trajectory planning algorithms: a general overview. in Motion and Operation Planning of Robotic Systems (G. Carbone, F. Gomez-Bravo, Eds.), *Springer*, Cham, 2015.
- [4]. P. Coufal, Š. Hubálovský, M. Hubálovská, Application of basic graph theory in autonomous motion of robots, *Mathematics*, Vol. 9, 2021, 919.
- [5]. A. Risald, et al., Best routes selection using Dijkstra and Floyd-Warshall algorithm, in *Proceedings of the* 11th International Conference on Information & Communication Technology and System (ICTS'17), Surabaya, Indonesia, 2017, pp. 155-158.
- [6]. C. S. Tan, R. Mohd-Mokhtar, M. R. Arshad, A comprehensive review of coverage path planning in robotics using classical and heuristic algorithms, *IEEE Access*, Vol. 9, 2021, pp. 119310-119342.
- [7]. P. C. Pop, O. Cosma, C. Sabo, C. Pop Sitar, A comprehensive survey on the generalized traveling salesman problem, *European Journal of Operational Research*, Vol. 314, Issue 3, 2024, pp. 819-835.

(023)

CFUs Detection in Petri Dish Images Using YOLOv12

V. Quevit^{1,2,3}, J.-L. Dillenseger¹, J.-M. Laferté², A.-J. Fougères², H. Djelal⁴ and E. Jalenques³

¹ Univ. Rennes, LTSI - UMR 1099, F-35000 Rennes, France
 ² ECAM Louis de Broglie, Campus de Ker Lann, Bruz, Rennes 35091, France
 ³ Interscience, 30 Chem. du Bois des Arpents, 78860 Saint-Nom-la-Bretèche, France
 ⁴ UniLaSalle Rennes - École des Métiers de l'Environnement, CYCLANN, Campus de Ker Lann, 35170 Bruz, France
 E-mail: victorien.quevit@ecam-rennes.fr

Summary: This study proposes to use the Deep Learning detection model, YOLOv12, for Colony-Forming Unit (CFU) detection in Petri dish images, aiming to automate the traditionally labor-intensive and error-prone manual counting process. YOLOv12 integrates attention mechanisms to enhance detection accuracy while maintaining real-time performance. The model achieves a mAP50 of 0.975 and a mAP50:95 of 0.706 across all 5 classes of CFU in the AGAR dataset, demonstrating its effectiveness in automating microbiological analysis. This innovation highlights the potential of YOLOv12 to streamline laboratory workflows and improve accuracy in CFU detection.

Keywords: Deep learning, Convolutional neural network, YOLO, Colony forming units, Petri dish.

1. Introduction

Due to the microscopic nature of microorganisms, which prevents them from being directly counted, Petri dishes serve as a controlled environment to facilitate their growth, resulting in the formation of Colony-Forming Units (CFUs) that are visible to the naked eye. This method, known as microbial culture [1], is fundamental in microbiology for quantifying and identifying microorganisms in various samples. The typical CFU count ranges from 30 to 300 on dishes with a standard diameter of 90 mm. Accurate enumeration of CFUs ensures the safety and quality of products across multiple industries, including food, cosmetics, and pharmaceuticals [2-5]. This process is not only a regulatory requirement but also a critical step in preventing microbial contamination, which can lead to health risks [6] (from foodborne illness to death), product recalls, and economic losses. Therefore, counting CFUs on Petri dishes is an indispensable procedure in quality control and safety process.

However, traditional methods for CFU counting predominantly rely on manual inspection by trained microbiologists. This approach is time-consuming, labor-intensive, and prone to human error, making it challenging to meet the growing demands for efficiency and accuracy in modern laboratories. Manual counting involves visually inspecting each Petri dish, identifying colonies, and recording their numbers, which can take several minutes per dish. This process becomes even more cumbersome when dealing with large-scale analyses, where hundreds or thousands of dishes need to be examined daily.

Additionally, the subjectivity of human judgment can lead to inconsistencies in results, further complicating the reliability of manual methods. To address these limitations, automated solutions have been developed, but many still rely on classical image processing techniques, such as thresholding, edge detection, and color segmentation. While these methods offer some level of automation, they often struggle with the complexity and variability of CFU appearances. Factors such as overlapping colonies, varying sizes, shapes, and colors, as well as the presence of artifacts like bubbles, writings or scratches on the Petri dish surface, can significantly impact the accuracy of these traditional algorithms. Furthermore, these methods often require extensive preprocessing and parameter tuning, which can be impractical for laboratories with diverse sample types and varying imaging conditions.

In recent years, advancements in artificial intelligence, particularly Deep Learning, have revolutionized the field of image analysis. Deep Learning models, such as Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs), have demonstrated exceptional performance in object detection, classification, and segmentation tasks [7]. While CNNs are faster and more efficient [8], they often fall short in capturing complex patterns and features compared to ViTs. Vision Transformers, on the other hand, excel in modeling intricate within relationships images [9] but are computationally and data intensive, complex, and less suited for real-time applications. This trade-off between speed and performance poses a significant challenge in the context of microbiological analysis, where rapid and accurate results are essential.

To address these challenges, we propose integrating YOLOv12 [10], a recently released (Feb. 24th 2025) state-of-the-art Deep Learning model designed for real-time object detection. YOLOv12 combines the strengths of both CNNs and ViTs,

offering a balance between speed and accuracy. By leveraging advanced attention mechanisms and optimized architectures, YOLOv12 achieves high detection performance while maintaining the efficiency required for real-time analysis. This makes it an ideal candidate for automating the detection and enumeration of CFUs in Petri dish images.

In this study, we will evaluate the performance of YOLOv12 in the context of microbiological image analysis to determine its effectiveness in meeting the demands of real-time CFU detection and enumeration.

Section 2 covers the methodology: Section 2.1 introduces the AGAR dataset, detailing its structure and relevance. Section 2.2 presents YOLOv12, highlighting key innovations like Area Attention and R-ELAN. Section 2.3 describes the training process, including dataset split, preprocessing, and model optimization. In Section 3, we compare YOLOv12m performance against Faster R-CNN [11], Cascade R-CNN, and Vision Transformers. Section 4 discusses future improvements, and Section 5 summarizes key findings.

2. Method

2.1. AGAR Dataset

For this study, we used the AGAR dataset by NeuroSYS [12] which is a comprehensive collection of microbial colony images. It consists of 18000 annotated images of Petri dishes, featuring five different microorganisms: Staphylococcus aureus, Pseudomonas Bacillus subtilis, aeruginosa, Escherichia coli, and Candida albicans (Table 1). These microorganisms were cultured under diverse lighting conditions and captured using two different cameras, resulting in images of varying resolutions. The dataset includes both single and mixed cultures, providing a rich and diverse set of samples for training and evaluating Deep Learning models.

The images in the AGAR dataset are categorized into two subsets based on resolution: higher-resolution (4000 \times 6000 pixels), lower-resolution (2048 \times 2048 pixels). For our study, we utilized the lower-resolution subset. This subset is particularly valuable for developing models that handle common real-world microbiological analysis scenarios and imaging conditions.

The annotations include bounding boxes (BBs) for each colony, specifying their location and class, which is essential for training object detection models.

In the associated paper, researchers evaluated two prominent object detection models, Faster R-CNN and Cascade R-CNN, to detect CFU in the AGAR dataset images. These models were chosen for their robust performance in object detection tasks and were trained and tested using the dataset. The results of these experiments will serve as a comparison benchmark for evaluating the performance of YOLOv12 in accurately identifying and counting CFUs in microbiological images.

Table 1. CFU Classes, Descriptions,	and Representations
from the AGAR Data	aset.

Classes	Description	Image (280 × 280)
S. aureus	A gram-positive bacterium commonly found in the human microbiota but also known for causing infections.	•
B. subtilis	A gram-positive bacterium widely used in industrial and scientific applications, known for its ability to form endospores.	
P. aeruginosa	A gram-negative bacterium known for its resistance to antibiotics and its role in hospital- acquired infections.	20: .
E. coli	A gram-negative bacterium commonly found in the human gut, with some strains causing foodborne illnesses.	
C. albicans	A yeast that is a common cause of fungal infections in humans, particularly in immunocompromi sed individuals.	•

2.2. YOLOv12 Model

YOLOv12 is designed to balance speed and precision, making it ideal for real-time applications, including microbiological analysis. Its architecture integrates advanced attention mechanisms, which enable it to focus on relevant features in an image while ignoring irrelevant details. This selective attention enables YOLOv12 to achieve a mean average precision (mAP) of 0.552 on the COCO 2017 dataset [13], a widely used benchmark for evaluating detection models, positioning it among the top-performing detection models.

At the core of YOLOv12 lies the attention-centric approach, which sets it apart from CNNs. While CNNs

excel in extracting spatial features, they often struggle with capturing long-range dependencies and complex relationships within an image. YOLOv12 addresses this limitation by incorporating attention mechanisms, which dynamically weigh the importance of different regions in an image. This allows the model to focus on the most informative parts of the image, such as the edges of microbial colonies or their distinctive textures, while disregarding less relevant areas like the background which represents most of the Petri dish images.

The three main improvements of YOLOv12 are:

Area Attention:

YOLOv12 introduces Area Attention, a novel mechanism that divides the image into vertical or horizontal segments, reducing computational complexity while maintaining a large receptive field. This approach ensures that the model can efficiently process high-resolution images, such as those in the AGAR dataset, without sacrificing accuracy. By focusing on specific areas of the image, Area Attention helps YOLOv12 to handle the variability in colony sizes, shapes, and densities, which are common challenges in microbiological analysis.

Residual Efficient Layer Aggregation Networks (R-ELAN):

YOLOv12 incorporates R-ELAN, an architecture designed to enhance feature aggregation and stabilize training. R-ELAN introduces residual connections and scaling techniques, which improve gradient flow and prevent issues like vanishing gradients during training.

Optimized Architecture:

YOLOv12 streamlines its architecture by removing redundant layers and optimizing the balance between convolutional and attention-based operations. This reduces computational overhead and improves inference speed. The model also leverages FlashAttention, a technique that optimizes memory access patterns during attention calculations, further enhancing efficiency.

The attention-centric approach in YOLOv12 is inspired by the success of transformer models in natural language processing and computer vision. Unlike traditional CNNs, which apply convolutional filters uniformly across an image, attention mechanisms allow the model to adaptively focus on specific regions. This is particularly useful in microbiological analysis, where colonies can vary widely in size, shape, and texture.

2.3. Models Training

The dataset (AGAR low-resolution subset) used for training consisted of 4237 images, while the validation set included 3475 images. Additionally, 2817 test images were used to evaluate the model performance. The images were originally captured at a resolution of 2048 \times 2048 pixels and were reduced to 1024 \times 1024 pixels to optimize resource usage during training.

The training of YOLOv12 was conducted over 50 epochs using a batch size of 6 and an image size of

1024 pixels on an NVIDIA RTX 4060 (16 GB) GPU. The model was trained from scratch, without pre-training, to ensure it learned directly from the provided dataset.

YOLOv12 is available in five versions: nano (n), small (s), medium (m), large (l), and extra-large (x). The model used was the medium version of YOLOv12, which offers a good compromise between resource consumption and performance, providing an excellent. This version operates at 67.5 GFLOPs (Giga Floating Point Operations Per Second), ensuring a good balance between performance and computational efficiency, making it suitable for real-time applications while respecting hardware limitations. This makes it an ideal choice for automated microbiological analysis, where both speed and accuracy are critical.

То enhance the model robustness and generalization capabilities, extensive data augmentation techniques were applied, including random flipping, rotation, zooming, HSV (hue, saturation, value) variations, and mosaic augmentation. These augmentations helped the model to better handle the variability in colony appearances and improved its performance on unseen data.

The training process lasted approximately 6 hours, with an estimated electricity consumption of 1.1 kWh, electricity cost of 0.20 € and a carbon footprint of 50 grams of CO , based on French energy standards. The model lightweight architecture and optimized training process contributed to minimal resource consumption, making it a cost-effective and environmentally friendly choice.

3. Results

Once YOLOv12m is trained on the AGAR lower-resolution subset, it can infer on unseen data from the test dataset. Below are a few examples of its inferences (Figs. 1, 2 and 3).



Fig. 1. YOLOv12m inference on a mixed-culture Petri dish image of *S. aureus* (blue BBs) and *P. aeruginosa* (white BBs). YOLOv12m handles low-contrast CFUs (*P. aeruginosa*) well.



Fig. 2. YOLOv12m inference on a mixed-culture Petri dish image of *S. aureus* (blue BBs) and *E. coli* (green BBs). YOLOv12m successfully detected *S. aureus* CFUs within *E. coli* CFUs.



Fig. 3. YOLOv12m inference on a single-culture Petri dish image of *B. subtilis* (light blue BBs), showcasing how well YOLOv12m detect overlapping CFUs.

The following table (Table 2) presents the performance metrics of YOLOv12m on the lower-resolution subset of the AGAR dataset. The evaluation includes precision, recall, mAP50, and mAP50:95, commonly used to evaluate object detection models [7].

 Table 2. Results obtained by YOLOv12m

 on the lower-resolution subset of the AGAR dataset.

Classes	Precision	Recall	mAP50	mAP50:95
S. aureus	0.971	0.919	0.970	0.685
B. subtilis	0.967	0.947	0.975	0.676
P. aeruginosa	0.970	0.976	0.989	0.738
E. coli	0.988	0.988	0.994	0.784
C. albicans	0.970	0.886	0.950	0.649
All	0.973	0.943	0.976	0.706

The model achieved high precision and recall scores for all classes, demonstrating its ability to accurately detect colonies while minimizing false positives and false negatives. Notably, *E. coli* showed the highest precision and recall, with values of 0.988, reflecting the model excellent performance in identifying this particular species. *P. aeruginosa* also exhibited strong performance, with a precision of 0.970 and a recall of 0.976. *S. aureus*, *B. subtilis*, and *C. albicans* similarly demonstrated high precision and recall. YOLOv12m achieved an overall mAP50 of 0.976 and mAP50:95 of 0.706. This indicates that the model not only accurately detects colonies but also maintains high precision across varying levels of overlap between predicted and ground truth BBs. Let's compare YOLOv12m with other CNNs models used in [11] trained with the AGAR dataset (Table 3).

Table 3. mAP50:95 results comparison between FasterR-CNN,CascadeR-CNNandYOLOv12on the lower-resolution subset of the AGAR dataset, bestresults in bold green.

Classes	Faster R-CNN	Cascade R-CNN	YOLOv12m
S. aureus	0.665	0.692	0.685
B. subtilis	0.441	0.480	0.676
P. aeruginosa	0.506	0.547	0.738
E. coli	0.565	0.582	0.784
C. albicans	0.652	0.668	0.649
All	0.560	0.594	0.706

The comparative results reveal significant insights into the performance of these models in detecting microbial colonies. YOLOv12m consistently outperformed Faster R-CNN and Cascade R-CNN for most of the classes.

Overall, YOLOv12m achieved an mAP50:95 of 0.706 across all classes, significantly outperforming both Faster R-CNN (0.560) and Cascade R-CNN (0.594). This superior performance can be attributed to YOLOv12 attention-centric approach, which allows it to focus on relevant features while ignoring irrelevant details. These results confirm YOLOv12 significant improvements over existing methods. But let's compare these results to those obtained with ViT models applied to the AGAR dataset in [14] (Table 4).

Table 4. mAP50:95 results comparison between Cascade Mask (CM) R-CNN, Mask R-CNN using Swin Transformer backbone and YOLOv12m on the AGAR dataset, best results in bold green.

Classes	CM R-CNN*		Mask R-CNN*	YOLOv12m
Backbone	Swin-B	Swin-S	Swin-T	
S. aureus	0.736	0.746	0.722	0.685
B. subtilis	0.531	0.520	0.508	0.676
P. aeruginosa	0.604	0.603	0.577	0.738
E. coli	0.426	0.422	0.406	0.784
C. albicans	0.772	0.753	0.752	0.649
All	0.614	0.609	0.560	0.706

* Please note that the Swin Transformer-based backbone models were trained using a different technique and subset on the AGAR dataset (as described in their paper [14]) but achieved similar results using this technique with Faster R-CNN and Cascade R-CNN (with no Swin backbone), as those shown in Table 2. YOLOv12m achieves an mAP50:95 of 0.706, outperforming both Cascade Mask R-CNN and Mask R-CNN model using Swin Transformer backbones. Meanwhile, YOLOv12m maintains a lower computational cost with 67.5 GFLOPs and 20.2 million parameters, making it well-suited for real-time applications. In contrast, Swin Transformer backbones offer competitive performance but at a higher computational cost.

For instance, the Swin-T backbone in Mask R-CNN [15] achieves an mAP50:95 of 0.560 with 103.85 GFLOPs and 47.39 million parameters, while the Swin-S backbone reaches an mAP50:95 of 0.609 with 576.35 GFLOPs and 105.74 million parameters. The Swin-B backbone further increases the computational demand with 613.78 GFLOPs and 143.77 million parameters, achieving an mAP50:95 of 0.614 (Table 5).

Table 5. Comparison of GFLOPs and number of parameters between Cascade Mask (CM) R-CNN, Mask R-CNN using Swin Transformer backbone, and YOLOv12m, lowest computational cost in bold green.

Models	GFLOPs	Parametres (M)
CM R-CNN Swin-S	576.35	105.74
CM R-CNN Swin-B	613.78	143.77
Mask R-CNN Swin-T	103.85	47.39
YOLOv12m	67.5	20.2

While Swin Transformer backbones deliver high accuracy, their significantly higher computational requirements make them less practical for real-time applications compared to YOLOv12m. This balance between better performance and better efficiency positions YOLOv12m as a superior choice for automated microbiological analysis, especially in resource-constrained environments.

4. Discussion

While this study demonstrates the superior performance of YOLOv12m in detecting and enumerating CFUs in Petri dish images, there are several areas for further exploration and improvement. Firstly, the study did not fully utilize the high-resolution subset nor the native lower-resolution $(2048 \times 2048 \text{ pixels})$ available in the AGAR dataset. Training and evaluating YOLOv12m on higher-resolution images could potentially enhance its detection accuracy, especially for smaller or more densely packed colonies. Additionally, the model was trained for only 50 epochs due to computational constraints. With more extensive resource computational resources, a longer training duration could further refine the model performance and generalization capabilities.

Furthermore, this study focused on the medium version of YOLOv12, which offers a balanced trade-off between computational efficiency and

detection accuracy. However, the extra-large version of YOLOv12(x), although more computationally intensive, has been shown to achieve even better results on the COCO 2017 dataset. Investing in the extra-large version could yield higher detection accuracy, making it a valuable avenue for future research. Additionally, comparative studies with other state-of-the-art models, such as YOLOv8x [16], which has been reported to achieve a mAP50:95 of 0.767 [14] in this task, would provide valuable insights into the strengths and limitations of YOLOv12.

5. Conclusion

The results of this study demonstrate that YOLOv12m significantly outperforms both traditional CNN-based models and Vision Transformer models in detecting and enumerating CFUs in Petri dish images. YOLOv12m achieved an impressive mAP50:95 of 0.706 on the lower-resolution subset of the AGAR dataset, surpassing the performance of Faster R-CNN (0.560) and Cascade R-CNN (0.594), as well as ViT models like Cascade Mask R-CNN and Mask R-CNN with Swin Transformer backbones. This superior performance can be attributed to YOLOv12 advanced attention-centric approach, which allows it to focus on relevant features while ignoring irrelevant details, and its optimized architecture, which balances speed and accuracy.

YOLOv12m ability to maintain high detection accuracy while keeping computational costs relatively low (67.5 GFLOPs and 20.2 million parameters) makes it an ideal choice for real-time applications. With an inference time of just 36 milliseconds per image (or 27 images per second) on an NVIDIA RTX 4060 (16 GB), which can be considered a low- to mid-range GPU, YOLOv12 meets the demands of real-time microbiological analysis, ensuring both speed and precision. In contrast, ViT models, while delivering competitive, yet lower, accuracy, come with significantly higher computational requirements, making them less practical for real-time use.

The attention mechanisms in YOLOv12 enable the model to efficiently process high-resolution images and handle the variability in colony sizes, shapes, and densities. This makes YOLOv12m robust and reliable for automated microbiological workflows, where accuracy and speed are critical. YOLOv12 represents a major advancement in automated microbiological analysis, offering a powerful and efficient solution for real-time CFU detection. Its superior performance, combined with its computational efficiency, positions YOLOv12 as a leading tool for ensuring the safety and quality of products across various industries.

References

[1]. G. Kapinusova, M. A. Lopez Marin, O. Uhlik, Reaching unreachables: Obstacles and successes of

microbial cultivation and their reasons, *Frontiers in Microbiology*, Vol. 14, 2023.

- [2]. L. Consuelo, et al., Microbiological contamination of surfaces in fish industry, *African Journal of Microbiology Research*, Vol. 8, 2014, pp. 425-431.
- [3]. A. Chatterjee, J. Abraham, Microbial contamination, prevention, and early detection in food industry, in Microbial Contamination and Food Degradation (Handbook of Food Bioengineering), *Academic Press*, 2018, pp. 21-47.
- [4]. E. Cunningham-Oakes, et al., Understanding the challenges of non-food industrial product contamination, *FEMS Microbiology Letters*, Vol. 366, Issue 23, 2019, fnaa010.
- [5]. T. T. B. C. Ribeiro, G. Costa, M. da Costa, Microbial contamination in industrial tofu, *Ciência Rural*, Vol. 47, 2017, 20160234.
- [6]. K. S. Almaary, Food-borne diseases and their impact on health, *Biosciences Biotechnology Research Asia*, Vol. 20, Issue 3, 2023, 3129.
- [7]. E. Arkin, et al., A survey: Object detection methods from CNN to transformer, *Multimedia Tools and Applications*, Vol. 82, Issue 14, 2023, pp. 21353-21383.
- [8]. K. O'Shea, R. Nash, An introduction to convolutional neural networks, arXiv preprint, 2015, arXiv:1511.08458.

- [9]. A. Dosovitskiy, et al., An image is worth 16x16 words: transformers for image recognition at scale, *arXiv preprint*, 2021, arXiv:2010.11929.
- [10]. Y. Tian, Q. Ye, D. Doermann, YOLOv12: attention-centric real-time object detectors, arXiv preprint, 2025, arXiv:2502.12524.
- [11]. S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: towards real-time object detection with region proposal networks, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 39, Issue 6, 1 June 2017, pp. 1137-1149.
- [12]. S. Majchrowska, et al., AGAR: A microbial colony dataset for deep learning detection, *arXiv preprint*, 2021, arXiv:2108.01234.
- [13]. T. Y. Lin, et al., Microsoft COCO: common objects in context, in *Proceedings of the European Conference on Computer Vision (ECCV'14)*, 2014, pp. 740-755.
- [14]. F. Yang, et al., Microbial colony detection based on deep learning, *Applied Sciences*, Vol. 13, Issue 19, 2023, 10568.
- [15]. K. He, et al., Mask R-CNN, in Proceedings of the IEEE International Conference on Computer Vision (ICCV'17), 2017, pp. 2920-2988.
- [16]. D. Reis, et al., Real-time flying object detection with YOLOv8, arXiv preprint, 2024, arXiv:2305.09972.
7th International Conference on Advances in Signal Processing and Artificial Intelligence (ASPAI' 2025), 8-10 April 2025, Innsbruck, Austria

(026)

A Reliable and Efficient Detection Pipeline for Rodent Ultrasonic Vocalizations

S. S. Anis¹⁻⁴, D. M. Kellis⁵, K. F. Kaigler^{5,6}, M. A. Wilson^{5,6} and <u>C. O'Reilly</u>¹⁻⁴

 ¹ Artificial Intelligence Institute, University of South Carolina, Columbia, SC, USA
 ² Institute for Mind and Brain, University of South Carolina, Columbia, SC, USA
 ³ Carolina Autism and Neurodevelopment Research Center, University of South Carolina, Columbia, SC, USA
 ⁴ Department of Computer Science and Engineering, University of South Carolina, Columbia, SC, USA
 ⁵ Department of Pharmacology, Physiology & Neuroscience, University of South Carolina School of Medicine, Columbia, SC, USA
 ⁶ Columbia VA Health Care System, Columbia, SC, USA E-mail: sanis@email.sc.edu; christian.oreilly@sc.edu

Summary: Analyzing ultrasonic vocalizations (USVs) is crucial for understanding rodents' affective states and social behaviors, but the manual analysis is time-consuming and prone to errors. Automated USV detection systems have been developed to address these challenges. Yet, these systems often rely on machine learning and fail to generalize effectively to new datasets. To tackle these shortcomings, we introduce ContourUSV, an efficient automated system for detecting USVs from audio recordings. Our pipeline includes spectrogram generation, cleaning, pre-processing, contour detection, post-processing, and evaluation against manual annotations. To ensure robustness and reliability, we compared ContourUSV with three state-of-the-art systems using an existing open-access USV dataset (USVSEG) and a second dataset we are releasing publicly along with this paper. On average, across the two datasets, ContourUSV outperformed the other three systems with a $1.51 \times$ improvement in precision, $1.17 \times$ in recall, $1.80 \times$ in F1 score, and $1.49 \times$ in specificity while achieving an average speed up of $117.07 \times$.

Keywords: Rodents, Ultrasonic vocalizations (USVs), Contour detection, Signal processing.

1. Introduction

Rodents, such as rats and mice, use ultrasonic vocalizations (USVs) as a form of communication in various behavioral contexts. These vocalizations, which occur at frequencies beyond the range of human hearing, have become an important tool for studying the emotional states and social behaviors of rodents [1]. For example, USVs are often emitted in response to stress, mating, or social interactions, providing valuable insights into the neural mechanisms underlying these behaviors. Understanding and analyzing USVs can thus shed light on how rodents communicate, how their behavior changes in response to different stimuli, and how these processes may relate to human neuropsychiatric disorders such as anxiety, depression, post-traumatic stress disorder (PTSD), and autism [1].

USVs are high-frequency sounds (20-120 kHz) emitted by rodents in various behavioral contexts (e.g., mating, aggression, and distress) [1]. Traditional manual analysis of USVs is time-consuming, laborintensive, and error-prone. Additionally, manual methods suffer from subjectivity, with different researchers potentially annotating the same data differently [2]. Automated systems can address these limitations by accelerating the analysis of large datasets, improving consistency, and minimizing the need for manual annotation. Several systems have been developed using signal processing and machine learning techniques [3-9]. However, these systems often fail to generalize to new datasets and sometimes require extensive manual intervention [10]. Moreover, training deep learning models can be time-consuming and resource-intensive, leaving a significant carbon footprint.

We novel automated propose а and energy-efficient approach for detecting rodent USVs using robust contour detection on spectrograms. We evaluated the reliability of our system on two open-access datasets and compared its performance to state-of-the-art methods (DeepSqueak [4], Joseph the mouse (JTM) [6], and USVSEG [9]). Through rigorous experimental analysis, the ContourUSV detection pipeline demonstrated robustness across different datasets, offering a reliable and scalable solution for large-scale USV analysis.

2. Related Works

Several automated systems have been developed to detect USVs. MUPET [3], an open-source software, uses signal processing techniques for rapid, unsupervised analysis of mouse USVs. It provides automated discovery and comparison of syllable types (i.e., call types) across strains and social conditions. It also incorporates noise removal and time-stamping features to facilitate behavioral analysis. DeepSqueak [4] automates USV detection and classification using deep neural networks, clustering, and supervised classification. JTM [6] proposes two alternative 7th International Conference on Advances in Signal Processing and Artificial Intelligence (ASPAI' 2025), 8-10 April 2025, Innsbruck, Austria

techniques to detect USVs: Morphological Geodesic Active Contour (GAC) [11] and Faster R-CNN [12]. While GAC offers a configurable, non-trainable approach, Faster R-CNN uses neural networks to learn from annotated data. HybridMouse [7] uses a combination of convolutional and recurrent neural networks for automatic USV identification and annotation, outperforming DeepSqueak in recall and F1 score metrics. VocalMat [5] provides tools for USV detection and classification, emphasizing user customization. It supports supervised and unsupervised methods but requires significant manual intervention. A-MUD [8] is an algorithm designed to detect mouse USVs automatically using the STx acoustic software. A-MUD achieves lower error rates than commercial software and accelerates USV detection 4 to 12 times compared to manual annotations. USVSEG [9] uses signal processing techniques to detect USV segments amid noise and track spectral peaks for syllables. The performance of this system has been demonstrated across several rodent species on an open-access dataset of the same name. BootSnap [10] can classify calls in syllables but does not propose a USV detection method, relying instead on existing detectors. The authors of BootSnap compared multiple detection methods, including DeepSqueak, USVSEG, A-MUD, and MUPET. They found A-MUD and USVSEG to have higher true positive rates and A-MUD to have lower false detection rates. A study of multiple deep-learning algorithms for neonatal murine USV detection demonstrated high performance on an open-access dataset of recordings [13]. However, this dataset does not include manual annotations. Therefore, we could not use it for comparative analysis. Despite these advances, the accuracy and reliability of these systems across datasets and experiments remain understudied. Further, most of these systems use advanced machine learning approaches for a task that might be solvable by lighter, more predictable signal processing approaches. Occam's Razor suggests it might be advantageous to look at computationally simpler and lighter approaches if they perform at least similarly.

3. Datasets

To assess the performance of our approach, we released an open-access dataset called USCMed and utilized another open-access dataset, USVSEG.

3.1. USCMed Dataset

The USCMed Dataset was collected at the University of South Carolina School of Medicine. This dataset involves a study designed to examine individual differences in rat fear conditioning and extinction. During this protocol, we recorded the USVs from male Long Evans rats (N = 27) exposed to the following four experimental conditions (Fig. 1; [14, 15]):

- Fear Acquisition: Three light foot shocks paired with 10-second 2 kHz tones and separated by a one-minute inter-stimulus interval were administered;
- <u>Contextual Fear:</u> Recording in the same environment (cage), but without tones or shocks;
- 3) <u>Cued Fear Extinction:</u> Twenty tones presented at one-minute intervals without co-occurring shocks, in a different environment;
- 4) <u>Extinction Recall:</u> Same as for 3), but two days later.

The audio was recorded at 250 kHz with UltraVox XT (Noldus Information Technology, Inc., Leesburg, VA). Each recording was manually annotated for the USV call start and stop times, frequency at max amplitude, and mean amplitude. These manual annotations served as the gold standard for evaluating our ContourUSV detector. Along with this paper, we released a subset of the USCMed dataset with the audio recordings (N = 27) and gold standard annotations for the Context trial. The USCMed dataset is openly available at https://doi.org/10.5281/zenodo.14211069.



Fig. 1. Paradigm used to collect the USCMed dataset (created with BioRender.com).

3.2. USVSEG Dataset

The USVSEG [9] dataset consists of recordings from mice (N = 20; C57BL/6J, BALB/c, and Shank2- adult males, and juvenile C57BL/6J mice), rats (N = 7; adult females, in distressing and pleasant contexts), and gerbils (N = 2), recorded at 250 kHz using a commercial condenser microphone and an A/D converter (UltraSoundGate, Avisoft Bioacoustics, Berlin, Germany; SpectoLibellus2D, Katou Acoustics Consultant Office, Kanagawa, Japan). The USVSEG dataset includes manual USV annotations and is openly available at https://doi.org/10.5281/ zenodo.3428023.

4. ContourUSV Detection Pipeline

This section describes the architecture of the ContourUSV pipeline as shown in Fig. 2. The code for ContourUSV is available on GitHub (https://github.com/lina-usc/contourusv).

4.1. Spectrogram Generation

The initial step in our pipeline involves generating spectrograms from raw audio recordings. These

spectrograms serve as the basis for the subsequent processing and detection. We first read the audio signals from the wav files (Fig. 3) using the Python SciPy library [16] and transform these signals into spectrograms (i.e., time-frequency representations) using the Short-Time Fourier Transform (STFT) implemented in MNE-Python [17]. The STFT was computed using a window size of 2,500 samples and a time step of 5 ms. To ensure the pipeline is not sensitive to differences in sample rate, we used the resample function from MNE-Python to resample signals to 250 kHz, if necessary. This function uses the same approach as SciPy, relying on the Fast Fourier Transform. We focused on the 15-115 kHz frequency range to isolate relevant signal components. The frequency resolution (100 Hz) is determined by the window size and the sampling frequency. The result is a 2D NumPy array representing the spectrogram data (time-frequency representation). Then, the NumPy array is passed to the spectrogram pre-processing and cleaning stage.



Fig. 2. Architecture of the ContourUSV detection pipeline.



Fig. 3. Example audio signal loaded from a.wav file (8s).

For visualization, we tested various Matplotlib [18] colormaps, and chose 'viridis' as it provided the best visual clarity for analyzing USVs (Fig. 4).



Fig. 4. Raw spectrogram from audio recording (8 s).

4.2. Pre-processing and Cleaning

The raw spectrogram data is further pre-processed to enhance the contrast between the USVs and the background noise. First, we apply a median filter from SciPy to the spectrogram data. Median filtering is used to reduce noise in the image while preserving edges. It replaces each pixel's value with the median value in its neighborhood. Then, we normalize the spectrogram using OpenCV [19] to scale the pixel intensity values of the filtered image to the [0, 255] range. The normalization function then casts the result to an 8-bit unsigned integer type, as standard for grayscale (single channel) images and required for OpenCV thresholding functions [18]. Next, we apply a thresholding approach to binarize the spectrogram. We chose Otsu's thresholding [20] instead of global thresholding to dynamically determine the optimal threshold for consistent binarization of spectrograms. This eliminates the need for a fixed threshold, enhancing robustness against noise and ensuring a clear separation of USVs from the background for reliable contour detection. To improve the visibility of the USVs, a contrast-limited adaptive histogram equalization (CLAHE) [21] is applied to the binary spectrograms. This technique enhances the local contrast of the images and makes the USV contours more prominent by applying an adaptive histogram equalization with a user-defined contrast limit to prevent over-amplification of noise. Finally, we apply morphological operations [18], specifically closing, which is a combination of dilation followed by erosion, which smooths object boundaries, fills small holes, and connects close objects in the spectrograms. This step ensures that the detected USV contours are continuous and well-defined as shown in Fig. 5.



Fig. 5. Cleaned and pre-processed spectrogram.

4.3. Contour Detection

The contour detection stage focuses on identifying and extracting the contours of USVs from the pre-processed spectrogram image data. This critical step allows for the precise localization of USVs in the time and frequency domains. In the pre-processing steps, applying CLAHE after the first binarization results in non-binary spectrograms. Otsu's thresholding is applied again to binarize the spectrograms, providing a clear separation between the USVs and the background and allowing for more reliable contour detection. Contours were then extracted from the thresholded binary spectrograms using the OpenCV findContours function. This function retrieves the external contours, representing the boundaries of the detected USVs. Fig. 6 illustrates an example of detected USVs.



Fig. 6. Detected contours and bounding boxes (green).

4.4. Post-processing Annotations

For each detected contour, a bounding box is computed using the OpenCV boundingRect function (see Fig. 6) and the time and frequency boundaries are calculated based on the position and dimensions of the bounding box relative to the image.

4.5. Evaluation

To assess the performance of ContourUSV, we compared the detected USVs against our gold standard (manual annotations) using the performance metrics shown in Fig. 7. First, we obtained both the predicted calls (system output) and actual calls (gold standard, obtained through manual annotations). Each annotation (manual or automated) included the start and end times of detected USVs. These annotations were used to create two sets of binary labels (predicted and actual) encoding the presence or absence of USVs for every time point in the audio signal. For this process, time windows specified in the annotations were mapped to sample indices of the audio files. To evaluate the system's performance, we calculated key metrics by comparing the predicted binary labels to the gold standard (actual) labels across all time points. True positives (TP), false positives (FP), false negatives (FN), and true negatives (TN) were defined as usual (Fig. 7, left panel). Using these values, we computed the precision, recall, F1 score, and specificity (Fig. 7, right panel). Finally, we aggregated the results across all recordings to compute the mean and standard deviation for each metric, providing an overall performance evaluation of the detection systems across various conditions and datasets. The statistical significance of performance differences between detectors was assessed with paired t-tests.

5. Results

This section presents the detection results for the USCMed and USVSEG datasets tested with ContourUSV, DeepSqueak, JTM, and USVSEG.



Fig. 7. Definition of the metrics used for assessing the performance of the different systems.

5.1. USCMed Dataset Results

Tables 1 and 2 present the results on the USCMed Dataset for the Fear Acquisition and the Context conditions, respectively. Bold is used to indicate statistical significance when comparing the performances of the best vs. second-best detectors. Table 3 and Table 4 show the paired t-test results comparing the best-performing model with the other models for Fear Acquisition and Context conditions, respectively. Statistical significance is defined at p < 0.05.

Table 1. USCMed dataset fear acquisition trial results
(Mean \pm SD).

Metric	DeepSqueak	JTM	ContourUSV	USVSEG
Precision	0.23 ± 0.15	0.77 ± 0.35	$\boldsymbol{0.86 \pm 0.23}$	0.42 ± 0.24
Recall	$\boldsymbol{0.97 \pm 0.05}$	0.59 ± 0.17	0.76 ± 0.23	0.89 ± 0.10
F1 Score	0.35 ± 0.19	0.62 ± 0.23	$\textbf{0.80} \pm \textbf{0.22}$	0.53 ± 0.26
Specificity	0.22 ± 0.14	0.94 ± 0.16	0.99 ± 0.01	0.71 ± 0.16

Table 2. USCMed dataset context trial results $(Mean \pm SD)$.

Metric	DeepSqueak	JTM	ContourUSV	USVSEG
Precision	0.17 ± 0.11	0.99 ± 0.01	$\boldsymbol{0.99 \pm 0.01}$	0.51 ± 0.17
Recall	$\boldsymbol{1.00\pm0.00}$	0.22 ± 0.09	0.87 ± 0.05	0.42 ± 0.10
F1 Score	0.28 ± 0.15	0.36 ± 0.12	$\textbf{0.93} \pm \textbf{0.02}$	0.44 ± 0.08
Specificity	0.21 ± 0.13	1.00 ± 0.00	1.00 ± 0.00	0.95 ± 0.01

The computation time was also recorded for each model on the USCMed dataset. ContourUSV performed detection in 359 seconds (i.e., 6 minutes) which is $7.26 \times$ faster than DeepSqueak's execution time of 2,607 seconds (i.e., more than 43 minutes), 226.89 \times faster than JTM's execution time of 81462 seconds (i.e., more than 22 hours), and 51.32 \times faster than USVSEG's execution time of 18426 seconds (i.e., more than 5 hours).

5.2. USVSEG Dataset Results

Table 5 presents the evaluation metrics for different species on the USVSEG dataset. The paired

t-test results for these comparisons are presented in Table 6 for all species.

 Table 3. Fear acquisition trial paired t-test results

 comparing the best-performing model with other models

 for each metric.

Metric (Best Model)	Comparison	T-value	P-value
Precision (ContourUSV)	DeepSqueak	17.80	1.04e ⁻¹⁵
	JTM	2.09	4.69e ⁻²
	USVSEG	11.64	1.37e ⁻¹¹
Recall (DeepSqueak)	ContourUSV	4.51	1.32e ⁻⁴
	JTM	11.34	2.37e ⁻¹¹
	USVSEG	4.28	2.42e ⁻⁴
F1 Score (ContourUSV)	DeepSqueak	12.98	1.31e ⁻¹²
	JTM	4.90	4.86e ⁻⁵
	USVSEG	7.37	1.02e ⁻⁷
Specificity (ContourUSV)	DeepSqueak	28.39	1.53e⁻²⁰
	JTM	1.66	1.09e ⁻¹
	USVSEG	9.23	1.57e⁻⁹

Table 4. Context trial paired t-test results comparing
the best-performing model with other models
for each metric.

Metric (Best Model)	Comparison	T-value	P-value
Precision (ContourUSV)	DeepSqueak	30.73	2.63e-18
	JTM	3.60	1.77e-03
	USVSEG	11.98	1.39e-10
Recall (DeepSqueak)	ContourUSV	-10.83	8.09e-10
	JTM	38.80	2.66e-20
	USVSEG	25.41	1.07e-16
F1 Score (ContourUSV)	DeepSqueak	22.53	1.10e-15
	JTM	25.10	1.36e-16
	USVSEG	32.26	1.01e-18
Specificity (ContourUSV)	DeepSqueak	26.08	6.46e-17
	JTM	-4.15	4.98e-04
	USVSEG	17.83	9.53e-14

Since ControurUSV has been designed and tested in the context of a fear acquisition protocol, we focused on detecting rat distress calls. To test if the lower performance of this system on the USVSEG dataset could be due to a failure to detect 50 kHz (positively valenced) calls, we additionally tested the detectors on only the 22 kHz calls from the USVSEG dataset. In this context, ContourUSV achieved an F1 Score of 0.96 ± 0.03 , while USVSEG, DeepSqueak, and JTM obtained F1-scores of 0.97 ± 0.01 , 0.40 ± 0.22 , and 0.55 ± 0.17 , respectively.

6. Discussion

Our primary goal was to develop an automated approach for detecting USVs in rodents that addresses the limitations of existing methods in terms of reliability, accuracy, simplicity, and computational demand. This focus on efficiency and simplicity is particularly important for applications that require real-time or embedded implementation. For instance, embedding such a detector in hardware could enable the *online* detection of USVs during an experiment, allowing researchers to dynamically modify experimental conditions based on rodent vocalizations. Such functionality could be especially beneficial in biofeedback applications, where immediate responses to vocalizations are crucial, such as in studies of social interaction or stress response. Additionally, a computationally lightweight system facilitates deployment in resource-constrained environments, such as portable devices for field studies. Since every second of recordings takes about 0.057 s to process with ContourUSV on a modern laptop (MacBook Pro), the computational efficiency of this approach would permit such real-time USV detection. Although we did not design ContourUSV for such a purpose, only minimal modifications should be required to adapt the code for such applications. However, we did not test ContourUSV's performance for online detection, a topic that would require additional research.

Table 5. ContourUSV evaluation metricson the USVSEG dataset.

Metric	DeepSqueak	JTM Gerbil	ContourUSV	USVSEG	
Precision	0.17 ± 0.00	0.97 ± 0.02	0.66 ± 0.17	0.72 ± 0.04	
Recall	0.99 ± 0.05	0.28 ± 0.21	0.97 ± 0.01	0.97 ± 0.03	
F1 Score	0.29 ± 0.01	0.42 ± 0.26	0.78 ± 0.12	0.82 ± 0.04	
Specificity	0.04 ± 0.02	1.00 ± 0.00	0.89 ± 0.08	0.92 ± 0.01	
		Mouse			
Precision	0.10 ± 0.10	0.62 ± 0.45	0.39 ± 0.27	$\textbf{0.73} \pm \textbf{0.14}$	
Recall	0.99 ± 0.02	0.16 ± 0.19	0.59 ± 0.17	0.91 ± 0.05	
F1 Score	0.18 ± 0.14	0.23 ± 0.25	0.42 ± 0.23	$\textbf{0.80} \pm \textbf{0.09}$	
Specificity	0.09 ± 0.09	0.69 ± 0.47	0.89 ± 0.08	0.97 ± 0.03	
		Rat			
Precision	0.24 ± 0.12	0.98 ± 0.02	0.84 ± 0.29	0.95 ± 0.03	
Recall	1.00 ± 0.00	0.24 ± 0.23	0.84 ± 0.17	0.90 ± 0.07	
F1 Score	0.37 ± 0.16	0.35 ± 0.28	0.81 ± 0.24	0.93 ± 0.05	
Specificity	0.06 ± 0.06	1.00 ± 0.00	0.96 ± 0.07	0.99 ± 0.01	
All Species					
Precision	0.14 ± 0.11	0.73 ± 0.41	0.52 ± 0.33	0.78 ± 0.15	
Recall	$\textbf{0.99} \pm \textbf{0.02}$	0.19 ± 0.20	0.68 ± 0.21	0.91 ± 0.06	
F1 Score	0.23 ± 0.17	0.27 ± 0.26	0.54 ± 0.28	0.83 ± 0.10	
Specificity	0.08 ± 0.08	0.79 ± 0.41	0.91 ± 0.08	0.97 ± 0.03	

 Table 6. Paired t-test results comparing the best-performing model with other models for each metric on the USVSEG dataset (all species).

Metric (Best Model)	Comparison	T-value	P-value
Precision (USVSEG)	DeepSqueak	-28.24	4.04e⁻²²
	JTM	-0.67	5.09e ⁻⁰¹
	ContourUSV	-5.20	1.62e⁻⁰⁵
Recall (DeepSqueak)	ContourUSV	8.25	5.67e ⁻⁰⁹
	JTM	22.12	2.85e ⁻¹⁹
	USVSEG	7.22	7.38e ⁻⁰⁸
F1 Score (USVSEG)	DeepSqueak	-26.06	3.54e ⁻²¹
	JTM	-13.79	5.29e ⁻¹⁴
	ContourUSV	-6.58	3.89e ⁻⁰⁷
Specificity (USVSEG)	DeepSqueak	-64.55	5.22e⁻³²
	JTM	-2.29	2.96e ⁻⁰²
	ContourUSV	-4.12	3.02e⁻⁰⁴

When we initiated this project, we also wanted this detector to be open-source and not require proprietary software (e.g., MATLAB). Furthermore, the limited availability of high-quality USV datasets with reliable gold standard annotations is one of the key challenges faced in our experiments as well as in this research field. Hence, we released a subset of our dataset (including the manual annotations) publicly to support USV detection benchmarking in future studies.

In our comparative analysis, ContourUSV had a high F1 score for both datasets, demonstrating strong reliability. Large performance differences between datasets highlight the significant effect that dataset properties have on the effectiveness of these systems. Since we used the USCMed Dataset for development, ContourUSV may have an unfair advantage over other systems on this dataset. The same is likely true for the evaluation of the USVSEG detector on the dataset of the same name. Nevertheless, ContourUSV shows superior reliability across datasets. Benchmarking against additional datasets would be required to corroborate this superior reliability.

Moreover, for the USCMed dataset, the results published in this paper are only for male rats on the fear acquisition and context trials. Thus, this dataset contains mostly 22 kHz call types. However, experiments with female rats and other experimental conditions within this data collection protocol, which contain a wider variety of call types (e.g., 50 kHz calls), are in progress, and results for these experiments will be published in future work. We are also investigating various denoising approaches, including single-channel decompositions. We hope such additions will allow the automatic removal of various noise sources and make ContourUSV even more reliable when processing recordings with unexpected sources of artifacts.

7. Conclusion

The ContourUSV detection method was designed to identify and localize USVs within spectrograms generated from audio recordings. This method employs a combination of preprocessing techniques, including median filtering, Otsu's thresholding, morphological operations, and contour extraction, to enhance and detect the contours of USVs. ContourUSV utilizes the OpenCV findContours function to accurately detect and annotate the temporal and frequency boundaries of each vocalization.

Our comparative analysis shows that ContourUSV consistently achieves higher mean F1 scores in the USCMed dataset. On the other hand, for the USVSEG dataset, ContourUSV performs as the second best after the USVSEG model. Since the development of ContourUSV utilized the USCMed dataset, which includes only male rats, its performance on the USVSEG dataset, comprising various rodent species, does not match that of the USVSEG model. Nevertheless, ContourUSV still outperforms both DeepSqueak and JTM in terms of F1 scores. The gap in recall and accuracy for ContourUSV is likely due to the lack of extensive noise reduction or filtering. On average, across the two datasets, ContourUSV outperformed the other three systems with a $1.51 \times$ improvement in precision, $1.17 \times$ in recall, $1.80 \times$ in F1

score, and $1.49 \times$ in specificity, while achieving an average speed up of $117.07 \times$.

Clustering and classifying calls for syntax analyses is an important area of future development for USV analyses. Without accurately detecting the calls first, these next stages of behavior analysis based on USVs cannot bring fruitful results. Thus, the ContourUSV detection pipeline serves an important role in advancing USV-based analysis of rodent behavior.

Acknowledgments

This work was supported by COR's startup package at USC, merit Award # I01 BX001374 to MAW from the United States Department of Veterans Affairs Biomedical Laboratory Research and Development Service (VA BLRD), and the USC VP for Research [ASPIRE II award to MAW; SPARC award to DMK]. The authors would also like to acknowledge the contributions of Sydnie L. Mick and Alicia N. Thomas to the collection of the USCMed dataset.

References

- [1]. M. Wöhr, R. K. Schwarting, Affective communication in rodents: ultrasonic vocalizations as a tool for research on emotion and motivation, *Cell and Tissue Research*, Vol. 354, Issue 1, 2013, pp. 81-97.
- [2]. M. Premoli, S. Pietropaolo, M. Wöhr, et al., Mouse and rat ultrasonic vocalizations in neuroscience and neuropharmacology: State of the art and future applications, *European Journal of Neuroscience*, Vol. 57, Issue 12, 2023, pp. 2062-2096.
- [3]. M. Van Segbroeck, A. T. Knoll, P. Levitt, S. Narayanan, MUPET – mouse ultrasonic profile extraction: a signal processing tool for rapid and unsupervised analysis of ultrasonic vocalizations, *Neuron*, Vol. 94, Issue 3, 2017, pp. 465-485.
- [4]. K. R. Coffey, R. E. Marx, J. F. Neumaier, Deepsqueak: a deep learning-based system for detection and analysis of ultrasonic vocalizations, *Neuropsychopharmacology*, Vol. 44, Issue 5, 2019, pp. 859-868.
- [5]. A. H. Fonseca, G. M. Santana, G. M. Bosque Ortiz, et al., Analysis of ultrasonic vocalizations from mice using computer vision and machine learning, *Elife*, Vol. 10, 2021, e59161.
- [6]. A. Kania, W. Ormaniec, D. Zhylko, et al., Joseph the MoUSE – Mouse ultrasonic sound explorer, *SoftwareX*, Vol. 25, 2024, 101606.
- [7]. Y. Goussha, K. Bar, S. Netser, et al., HybridMouse: a hybrid convolutional-recurrent neural network-based model for identification of mouse ultrasonic vocalizations, *Frontiers in Behavioral Neuroscience*, Vol. 15, 2022, 810590.
- [8]. S. M. Zala, D. Reitschmidt, A. Noll, et al., Automatic mouse ultrasound detector (A-MUD): A new tool for processing rodent vocalizations, *PloS ONE*, Vol. 12, Issue 7, 2017, e0181200.
- [9]. R. O. Tachibana, K. Kanno, S. Okabe, USVSEG: A robust method for segmentation of ultrasonic

7th International Conference on Advances in Signal Processing and Artificial Intelligence (ASPAI' 2025), 8-10 April 2025, Innsbruck, Austria

vocalizations in rodents, *PloS ONE*, Vol. 15, Issue 2, 2020, e0228907.

- [10]. R. Abbasi, P. Balazs, M. A. Marconi, et al., Capturing the songs of mice with an improved detection and classification method for ultrasonic vocalizations (BootSnap), *PLoS Computational Biology*, Vol. 18, Issue 5, 2022, e1010049.
- [11]. P. Marquez-Neila, L. Baumela, L. Alvarez, A morphological approach to curvature-based evolution of curves and surfaces, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 36, Issue 1, 2013, pp. 2-17.
- [12]. S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: Towards real-time object detection with region proposal networks, in Advances in Neural Information Processing Systems, Vol. 28, *Curran Associates, Inc.*, 2015.
- [13]. R. Herdt, L. Kinzel, J. G. Maaß, et al., Enhancing the analysis of murine neonatal ultrasonic vocalizations: Development, evaluation, and application of different mathematical models, *The Journal of the Acoustical Society of America*, Vol. 156, Issue 4, 2024, pp. 2448-2466.
- [14]. D. M. Kellis, K. F. Kaigler, E. Witherspoon, et al., Cholinergic neurotransmission in the basolateral amygdala during cued fear extinction, *Neurobiology of Stress*, Vol. 13, 2020, 100279.

- [15]. S. C. Tryon, I. M. Sakamoto, K. F. Kaigler, et al., ChAT::CRE transgenic rats show sex dependent altered fear behaviors, ultrasonic vocalizations and cholinergic marker expression, *Genes, Brain and Behavior*, Vol. 22, Issue 1, 2023, e12837.
- [16]. P. Virtanen, R. Gommers, T. E. Oliphant, et al., SciPy 1.0: fundamental algorithms for scientific computing in Python, *Nature Methods*, Vol. 17, Issue 3, 2020, pp. 261-272.
- [17]. A. Gramfort, M. Luessi, E. Larson, et al., MEG and EEG data analysis with MNE-Python, *Frontiers in Neuroinformatics*, Vol. 7, 2013, 267.
- [18]. J. D. Hunter, Matplotlib: A 2D graphics environment, *Computing in Science Engineering*, Vol. 9, Issue 03, 2007, pp. 90-95.
- [19]. G. Bradski, The OpenCV library, Dr. Dobb's Journal: Software Tools for the Professional Programmer, Vol. 25, Issue 11, 2000, pp. 120-123.
- [20]. N. Otsu, A threshold selection method from gray-level histograms, *Automatica*, Vol. 11, Issue 285-296, 1975, pp. 23-27.
- [21]. K. J. Zuiderveld, Contrast limited adaptive histogram equalization, *Graphics Gems*, Vol. 4, Issue 1, 1994, pp. 474-485.

(027)

Functional Connectivity Analysis Using Adaptive Window Size and Intersection of Confidence Intervals

Z. Šverko¹, <u>S. Vlahinić</u>², N. Stojković¹ and P. Rogelj³

 ¹ University of Rijeka, Faculty of Engineering, Department of Electric Power Systems, Vukovarska 58, 51000 Rijeka, Croatia
 ² University of Rijeka, Faculty of Engineering, Department of Automation and Electronics, Vukovarska 58, 51000 Rijeka, Croatia
 ³ University of Primorska, Faculty of Mathematics, Natural Sciences and Information Technologies, Glagoljaška 8, 6000 Koper, Slovenia Tel.: +385 51 505720

E-mail: zoran.sverko@riteh.uniri.hr, sasa.vlahinic@riteh.uniri.hr, nino.stojkovic@riteh.uniri.hr,

peter.rogelj@upr.si

Summary: This paper introduces a novel method for analyzing functional connectivity employing the absolute value of the complex Pearson correlation coefficient and the relative Intersection of Confidence Intervals algorithm. The method adapts window sizes based on signal variability and the window size is used for connectivity assessment. The approach was validated using synthetic *EEG* signals generated via the Kuramoto model, which ensured a realistic representation of connectivity dynamics. Additionally, the method was tested on real *EEG* data to evaluate its practical applicability. Results demonstrated the potential for differentiating low and high connectivity cases, with clear correlations between window size and statistical properties of phase differences. The findings highlight the potential of this adaptive methodology to provide more accurate and meaningful insights into functional connectivity, especially in dynamic systems where traditional fixed-window approaches fall short.

Keywords: EEG signal, Functional connectivity analysis, Adaptive window size.

1. Introduction

EEG is an electrophysiological technique used to observe neurophysiological changes related to postsynaptic activity in the neocortex, essentially recording the brain's electrical activity [1]. Brain connectivity analysis is generally categorized into two types: structural and functional. Structural connectivity analysis involves tracking the direction of fibers between different regions of the brain or within a specific region [2]. Functional connectivity analysis, on the other hand, examines the amount of information transferred between brain regions or within a single region [3]. This type of analysis is commonly split into two categories: undirected and directed. Undirected connectivity measures assess the strength of connectivity, whereas directed connectivity measures evaluate both the strength and direction of connectivity between the regions of interest. Various approaches can be employed to assess functional connectivity, phase synchronization, including generalized synchronization metrics, linear temporal correlation, and others [4-6]. In this paper, we concentrate on undirected connectivity measures. The phase locking value (PLV) [7, 8] and the weighted phase lag index (wPLI) [9] are among the most frequently used undirected phase synchronization metrics. The main difference between these two metrics is their ability to reduce the effects of volume conduction. In [3], it is demonstrated how the absolute and imaginary components of the complex Pearson correlation

coefficient (*CPCC*) exhibit properties similar to those of the *PLV* and *wPLI* metrics.

In this paper, a new measure for phase connectivity is proposed. It is based on dynamic connectivity analysis using absolute value of the complex Pearson correlation coefficient (*absCPCC*) as the connectivity measure and the relative Intersection of Confidence Intervals (*RICI*) algorithm for determining the window size [10]. The window size obtained in this way is used to assess the connectivity strength.

The development of a new measure is presented and applied to both synthetic and real *EEG* signals. It is also applicable to other methods of measuring brain activity, such as functional magnetic resonance imaging (*fMRI*) [11], blood-oxygen-level dependent (*BOLD*) signals [12], magnetoencephalography (*MEG*) signals [13], and others.

2. Methods

The *RICI* algorithm estimates the optimal window size based on the statistical properties of the signal, specifically through confidence intervals (hence the name, Intersection of Confidence Intervals) [10]. It relies on calculating the ratio between a quantitative criterion (R) and a threshold value (R_c), which serves to detect significant shifts in connectivity. For each time point, an initial window size is selected, which defines the duration of the temporal segment for analysis. For each window, a confidence interval (*CI*)

is computed, representing the range within which the true value of functional connectivity is expected to lie. Based on the computed *CI*, the upper and lower boundaries are determined. The minimal and maximal values of the upper and lower boundaries are updated based on the new data, ensuring accuracy in the analysis. With the updated boundary values, the ratio *R* is recalculated. When $R > R_c$, the process stops, and the last window for which this condition was not met is used to calculate the functional connectivity (*absCPCC*) in that window for the observed sample [3, 4]. *absCPCC* is defined as:

$$absCPCC = \frac{\left|\sum_{n=1}^{N} A_{x_{1,n}} \cdot A_{x_{2,n}} \cdot e^{j\Delta\varphi_{x_{1,n}x_{2,n}}}\right|}{\sqrt{\sum_{n=1}^{N} A_{x_{1,n}}^{2}} \cdot \sqrt{\sum_{n=1}^{N} A_{x_{2,n}}^{2}}},$$
(1)

where $\Delta \phi_{x_{1,n}x_{2,n}}$ is the instantaneous phase difference, A_x is the instantaneous amplitude of a complex signal.

When there is a stable phase difference, indicating high connectivity, the *RICI* algorithm identifies low variability interval and extends the window size as much as possible, up to the point where there is a significant change in the statistical properties of the signal. Conversely, when phase connectivity is low, the statistical properties of the phase differences fluctuate frequently, resulting in a narrow window size.

The proposed measure of phase connectivity is the average window size determined by the *RICI* algorithm applied to the dynamically calculated *absCPCC*.

3. Results

3.1. Synthetic Signals

The signals used in this study are synthetically generated using the Kuramoto model. Each signal is a mixture of two components. The first component of the signal is synchronized within the signal group, ensuring coherence, while the second component is arbitrary, introducing randomness that makes the signals more realistic and representative of natural signal dynamics [14]. The proposed method for phase connectivity analysis was applied to pairs of synthetically generated signals with varying connectivity strengths.

Fig. 1, illustrates the results of connectivity analysis using different approaches. Phase differences were analyzed within a 10 second window. *PLV* below 0.11 and *absCPCC* of 0.012 indicates low connectivity. The histogram of phase differences is presented in Fig. 1a, where the fairly uniform phase difference distribution confirms low connectivity for this pair of electrode.

The time evolution of phase differences, shown in Fig. 1b, reflects the characteristics of the model used

for generating the synthetic signals. The time diagram indicates the absence of periods with stable phase differences.



Fig. 1. Connectivity analyses for an electrode pair with a low connectivity, a) histogram of the phase differences in a 10 s window, b) time diagram of the same pair of signals, c) window size variations due to different statistical properties of the phase differences.

Finally, Fig. 1c presents the time diagram of the window size. The relatively low window size is attributed to the low phase connectivity of the signals and the high variability in the statistical properties of the phase differences. The *RICI* algorithm maintains a narrow window size. The maximum value of the window size was set to 2560 samples, corresponding to the observation period. The average relative window size was 0.10, further confirming the low connectivity, as expected.

Different behavior of the connectivity measure is anticipated in cases of high connectivity. Connectivity analysis for a synthetically generated signal pair with high phase connectivity is shown in Fig. 2. In Fig. 2a, phase differences are highly concentrated around a major peak at 1 radian and a minor peak at approximately -2.19 radians. The time diagram of the phase differences is shown in Fig. 2b. The two signals are phase-synchronized for longer periods around 1 radian and shorter periods around -2.19 radians.



Fig. 2. Connectivity analyses for an electrode pair with a high connectivity, a) histogram of the phase differences in a 10 s window, b) time diagram of a phase difference of the same pair of signals, c) window size variations due to different statistical properties of the phase differences.

In this case also, the window size is an effective indicator of the stable statistical properties of phase differences. The window size aligns with the periods of stable phase synchronization between the signals. During intervals of shorter phase synchronization, the statistical variations are higher, resulting in a lower window size maintained by the RICI algorithm.

Pairs with varying phase connectivity were generated, and this pair had the largest window size of 2258 samples and the lowest window size 154 samples. The average relative window size was 0.88, while *PLV* and *absCPCC* were 0.74 and 0.51, respectively, confirming the high connectivity.

3.2. Real Signals

For testing the proposed connectivity measure on real signals, the "SPIS Resting State Dataset" [15] was used. This dataset is multimodal and consists of *EEG* signals as well as electrooculogram (*EOG*) signals. For the purpose of evaluating the proposed static functional connectivity measure, only 10 seconds of *EEG* signals recorded in the eyes-closed (*EC*) condition for the participant "S02_restingPre_EC" were used (ensuring consistency with the duration of the synthetic signal, an interval of 15 to 25 seconds is selected), with a sampling frequency of 256 Hz.

Fig. 3 presents the outcomes of connectivity assessment using different methodologies. Phase differences were evaluated over a 10-second time window. A *PLV* value of 0.075 and an *absCPCC* of 0.023 indicate a weak level of connectivity. In Fig. 3a, the histogram of phase differences displays a relatively uniform distribution, reinforcing the notion of low connectivity between this particular electrode pair.



Fig. 3. Connectivity analyses for an electrode pair (P8-CP2) with a low connectivity, a) histogram of the phase differences in a 10 s window, b) time diagram of the same pair of signals, c) window size variations due to different statistical properties of the phase differences.

The temporal progression of phase differences, illustrated in Fig. 3b, shows fluctuations over time. The absence of prolonged periods with stable phase differences further supports the weak connectivity.

Fig. 3c shows the evolution of the window size over time. The relatively small window size results from the high variability in the statistical properties of the phase differences and the limited phase coupling of the signals. The *RICI* algorithm dynamically adapts the window size, keeping it relatively narrow. The average relative window size was recorded at 0.25, estimating the low connectivity.

A contrasting scenario is depicted in Fig. 4, which represents an instance of high phase connectivity. In Fig. 4a, the phase differences are predominantly concentrated around a primary peak, highlighting the presence of strong synchronization between the signals.



Fig. 4. Connectivity analyses for an electrode pair (P8-P4) with a high connectivity, a) histogram of the phase differences in a 10 s window, b) time diagram of a phase difference of the same pair of signals, c) window size variations due to different statistical properties of the phase differences.

The temporal representation of phase differences in Fig. 4b demonstrates extended intervals of stable phase relationships, indicating significant connectivity.

The evolution of the window size, displayed in Fig. 4c, correlates with these stable synchronization periods. The *RICI* algorithm effectively adjusts the window size to match the variations in statistical stability. The largest observed window size was 2560 samples, while the smallest was considerably reduced during phases of transient connectivity. The average relative window size of 0.94, alongside *PLV* and *absCPCC* values of 0.72 and 0.79, respectively, substantiates the strong connectivity between the signals.

4. Conclusions

The proposed methodology offers alternative approach to functional connectivity estimation. By dynamically adjusting the window size, using the *RICI* algorithm, this method successfully adapts to signal variability. The method was applied to synthetic and real signals with both high and low connectivity. The results confirm its potential in distinguishing between varying levels of connectivity, as demonstrated through synthetic and real *EEG* signal testing. This adaptive approach holds promise for applications in neuroscience and other fields requiring precise connectivity assessments. Future work will explore real-world *fMRI*, *BOLD*, *MEG* datasets.

Acknowledgements

This work has been supported by the University of Rijeka under the project number UNIRI-ISKUSNITEHNIC-23-31.

Z. Šverko: "I would like to thank the ERASMUS+ organization for the mobility scholarship, number of project: 2024-1-HR01-KA131-HED-000197229, during which this work was created."

References

- S. Sanei, J. A. Chambers, EEG Signal Processing, 1st Ed., *John Wiley & Sons*, Hoboken, NJ, USA, 2016, pp. 1-8.
- [2]. M. A. Koch, D. G. Norris, M. Hund-Georgiadis, An investigation of functional and anatomical connectivity using magnetic resonance imaging, *Neuroimage*, Vol. 16, 2002, pp. 241-250.
- [3]. Z. Šverko, M. Vrankić, S. Vlahinić, P. Rogelj, Complex Pearson correlation coefficient for EEG connectivity analysis, *Sensors*, Vol. 22, Issue 4, 2022, 1477.
- [4]. J. Sun, X. Hong, S. Tong, Phase synchronization analysis of EEG signals: an evaluation based on surrogate tests, *IEEE Transactions on Biomedical Engineering*, Vol. 59, Issue 8, 2012, pp. 2254-2263.
- [5]. C. J. Stam, B. W. Van Dijk, Synchronization likelihood: an unbiased measure of generalized synchronization in multivariate data sets, *Physica D: Nonlinear Phenomena*, Vol. 163, Issue 3-4, 2002, pp. 236-251.
- [6]. X. Bornas, M. Noguera, M. Balle, et al., Long-range temporal correlations in resting EEG, *Journal of Psychophysiology*, Vol. 27, Issue 2, 2013, pp. 60-66.
- [7]. C. J. Stam, G. Nolte, A. Daffertshofer, Phase lag index: assessment of functional connectivity from multi

channel EEG and MEG with diminished bias from common sources, *Human Brain Mapping*, Vol. 28, Issue 11, 2007, pp. 1178-1193.

- [8]. J. P. Lachaux, E. Rodriguez, J. Martinerie, F. J. Varela, Measuring phase synchrony in brain signals, *Human Brain Mapping*, Vol. 8, Issue 4, 1999, pp. 194-208.
- [9]. M. Vinck, R. Oostenveld, M. Van Wingerden, et al., An improved index of phase-synchronization for electrophysiological data in the presence of volume-conduction, noise and sample-size bias, *Neuroimage*, Vol. 55, Issue 4, 2011, pp. 1548-1565.
- [10]. Z. Šverko, M. Vrankić, S. Vlahinić, P. Rogelj, Dynamic connectivity analysis using adaptive window size, *Sensors*, Vol. 22, Issue 14, 2022, 5162.
- [11]. J. Hlinka, M. Paluš, M. Vejmelka, D. Mantini, M. Corbetta, Functional connectivity in resting-state fMRI: is linear correlation sufficient?, *Neuroimage*, Vol. 54, Issue 3, 2011, pp. 2218-2225.
- [12]. G. Pfurtscheller, A. Schwerdtfeger, D. Fink, et al., MRI-related anxiety in healthy individuals, intrinsic BOLD oscillations at 0.1 Hz in precentral gyrus and insula, and heart rate variability in low frequency bands, *PLoS One*, Vol. 13, Issue 11, 2018, e0206675.
- [13]. F. J. Hsiao, Z. A. Wu, L. T. Ho, Y. Y. Lin, Theta oscillation during auditory change detection: an MEG study, *Biological Psychology*, Vol. 81, Issue 1, 2009, pp. 58-66.
- [14]. Z Šverko, J. Sajovic, G. Drevenšek, S. Vlahinić, P. Rogelj, Generation of oscillatory synthetic signal simulating brain network dynamics, in *Proceedings of* the 44th International Convention on Information, Communication and Electronic Technology (MIPRO'21), 2021 Sep 27, pp. 141-146.
- [15]. M. Torkamani-Azar, S. D. Kanik, S. Aydin, M. Cetin, Prediction of reaction time and vigilance variability from spatio-spectral features of resting-state EEG in a long sustained attention task, *IEEE Journal of Biomedical and Health Informatics*, Vol. 24, Issue 9, 2020, pp. 2550-2558.

(028)

Chart Pattern Recognition Using Convolutional Neural Networks

C. Caballero-Gil^{1,} J. A. Antúnez-Pulido¹ and J. Giner-Rubio²

¹Department of Computer Engineering and Systems, University of la Laguna, Tenerife, Spain ²Department of Economics, Accounting and Finance, University of la Laguna, Tenerife, Spain E -mail: ccabgil@ull.edu.es, alu0101441702@ull.edu.es, jginer@ull.edu.es

Summary: In this study, convolutional neural network (CNN) technologies were implemented to detect chart patterns in stock market data. To train the CNN, a dataset comprising chart pattern images was developed, utilizing a synthetic pattern generation script and implemented in a program coded in Python. The resulting tool enables the identification of various chart pattern types across a specified set of companies and time periods. Beyond identifying patterns, the tool assesses whether the detected patterns fulfill their intended objectives. This capability allows the tool to compute the success rate of a given pattern over a defined period and for the selected companies. Additionally, the tool can identify whether a pattern is currently forming.

Keywords: Convolutional neural network, Technical analysis, Stock market, Chartist patterns.

1. Introduction

This paper presents a sophisticated tool to assist investors in their decision-making. One of the principles of technical analysis is that market movements are cyclical, so, as preached by technical analysis [1], knowing the past can help us to anticipate the future. The use of chart patterns is one of the most widely employed technical analysis techniques by traders. However, it is not the only one in technical analysis, as there are various other methods such as moving averages, Bollinger Bands, and many others, which are beyond the scope of this work.

The implemented tool will provide the number of patterns formed in a period for a subset of companies, and with the exact percentage of success or failure of each of the analyzed chart patterns. This approach enables investors to refine their investment strategies by analyzing specific time periods and subsets of companies, such as those within a targeted sector. By identifying the most effective chart patterns and their success rates, investors can make more informed decisions. Additionally, the tool facilitates the detection of emerging patterns in the analyzed companies, providing valuable insights for timely investment opportunities.

This work will use CNN (Convolutional Neural Network) to be chosen over a TCN (Temporal Convolutional Network) for training the detection of chart patterns in stock markets due to its ability to effectively capture spatial hierarchies and local patterns in visual data. Since chart patterns are often represented in image-like formats, CNNs are well-suited for detecting these patterns due to their strength in processing grid-like data. TCNs, on the other hand, are more specialized for handling sequential or temporal data with long-range dependencies, which, while useful for time-series forecasting, were less relevant in this context where local spatial patterns are more crucial for accurate pattern recognition. The tool has been implemented using Python, and the handling of large volumes of data through specialized libraries such as 'Pandas'. In addition, we have focused on the interpretation and presentation of the results in an accessible and useful way for the end user. Ultimately, it will allow for precise determination of the number of chart patterns identified for a set of companies, as well as the success rate of these patterns.

1.1. Patterns in Chart Analysis

Within technical analysis there is chartist analysis, which studies the shape of price charts to predict the future trend. These repeating patterns are chart structures that allow us to know, with some degree of certainty, future prices. Chartist analysis focuses on identifying periodic repetitions in the time series, non-linear patterns that can help us anticipate market movement.

According to the study by Farias et al. [2], chartist analysis was only used in 3 of the 85 relevant articles analyzed. For example, Dawson and Steeley [3] analyzed the existence of repetitive patterns in the UK stock market. More generally, Friesen et al. [4] also devote their efforts to the study of chartist techniques.

For this work, the search of chart patterns such as double top, double bottom, ascending triangle, descending triangle, head and shoulders, and inverted head and shoulder for a given period of dates in the selected companies will be allowed, and the tool will also calculate the percentage of companies that reach the objective with the patterns found.

1.2. Data Creation for Neural Network Training

The creation of the training database consisted of two phases. First, a script was created to search for historical patterns and save the image and a.csv file with the pattern points in a directory. The pattern search was conducted visually, supported by specialized websites such as Finviz, and a previously developed tool that uses the dynamic time warping technique for pattern detection, enabling the automatic identification of such patterns. Then, by hand, the highest quality patterns of those found were chosen and a script was programmed to create synthetic patterns based on a pattern passed by a parameter.

Data augmentation through the generation of synthetic images enhances the diversity of training data, particularly interesting when real data is scarce or difficult to obtain. In this paper, it involved mainly the scaling. By doing so, it helps prevent overfitting and improves the network's generalization ability. Additionally, synthetic images can be used to create cases that are not easily found such as some type of chart patterns. In summary, synthetic images expand the training dataset, improve model accuracy and robustness, and address challenges related to limited or hard-to-acquire data.

To train this model, a laptop (GF63 Thin 11UC) with an NVIDIA® GeForce RTX[™] 3050 graphics card and limited capacity for training models was used. The PyTorch library was employed, and two models were trained: one with a network for each pattern and a general network for all patterns. The architecture for both is the same, with only the number of neurons in the output layer differing. There are two output neurons in the case of a general network corresponding to the six types of patterns in the tool and the 'no pattern' class. These training sessions were carried out at night and in no case took more than 8 hours. It was used 2300 images per pattern type for the training.

Once the network development was completed, the training process began, with the data divided into training and testing sets, using 80 % for training and the remaining 20 % for testing. The training set was used to adjust the model parameters, while the test set, which has not been seen during training, is used to evaluate how the model generalizes new and independent data, allowing us to detect problems such as overfitting, where the model performs well on the training data but fails on unseen data. At the end of the training, a file with the model parameters is saved for later use. For both the pattern-specific networks and the general network, the accuracy on the testing set ranged from 96 % to 98 %, depending on the pattern.

Chart pattern classification labels typically refer to the categories or types of patterns identified in price charts. These labels include continuation patterns, which suggest the continuation of an existing trend (e.g., triangles, flags, wedges); reversal patterns, which indicate a potential trend reversal (e.g., head and shoulders, double top, double bottom); and consolidation patterns, which represent market indecision with sideways price movement (e.g., rectangles, symmetrical triangles).

To obtain synthetic patterns, new points were added to the original pattern, based on the average between the two points between which they were introduced, and existing points were modified by subtracting or adding between 10 % and 50 % of their value. The number of points to be added is a random value between 30 % and 50 % of the original number of points, while the number of points to be modified is a random value between 40 % and 70 %. These percentages were determined empirically. Excessively conservative percentage adjustments result in synthetic patterns that closely resemble the original, whereas overly aggressive adjustments can cause the pattern to lose its structural integrity. Fig. 1 illustrates an example of a run in which three synthetic patterns were generated based on a head and shoulders pattern.



Fig. 1. Original pattern and 3 synthetic patterns based on this pattern.

The network consists of convolutional layers that extract local features from the image, detecting patterns such as edges, textures, and basic shapes at different levels of abstraction. Each convolutional layer applies filters to the input image and generates feature maps that highlight certain visual properties. The output of the convolutional layer is passed to a pooling layer that reduces the dimensionality of the feature maps, reducing the number of parameters and the computational load, as well as providing some translation invariance (small variations of the image). Finally, the output of the last pooling layer is flattened into a feature vector and linear layers are applied that perform the final combination and transformation of the features extracted by the convolutional layers, allowing the network to make classification decisions based on the combined features learned from the images. Training the network consisted of adjusting the model parameters (weights and biases) by backpropagation and iterative optimization, using a training data set. During this process, the model makes predictions on the training data, calculates the loss or error by comparing the predictions with the actual labels, and then adjusts the parameters to minimize this loss using the stochastic gradient descent algorithm. The first and second steps were automatic. In training, functions were developed that converted window to image and image to tensor. In the case of a pattern network, the function that performs the classification receives a list of models (one model for each type of pattern) and each model performs the classification, which returns the predicted class (whether it is the pattern) and a confidence value, at the end it keeps the class with the highest confidence. For the joint network, it is simpler, only the classification must be performed, and it returns the predicted class.

The pattern network takes on average 50 % longer but finds on average 20 % more patterns, so we opted

for this one. Finally, a study was conducted to compare the search for historical patterns using a neural network. The study presented in this paper used 10 companies from the financial sector and 10 companies from the technological sector. The financial companies were V, JPM, BLK, SAN, BBVA, BBD, DB, NDAQ, BBAR, FCF. The technological companies were AAPL, MSFT, NVDA, ORCL, CSCO, NOK, WDC, LOGI, LPL, NEON. Fig. 2 presents examples of identified patterns and their alignment with predefined objectives. It illustrates a head-and-shoulders pattern successfully achieving its target, contrasted with another that fails to do so. Additionally, an inverted head-and-shoulders pattern meets the target, whereas ascending and descending triangles do not. The figure also includes a double-top pattern that successfully fulfills the target criteria.



Fig. 2. Detection of chart patterns using the implemented tool. Some identified patterns meet the expected objective, while others do not.

Table 1 shows the number of chart patterns found for the testing including financial and technological companies from January 1, 2005 until July 3, 2023.

4. Conclusions

In this work, a tool for pattern detection using convolutional neural networks for pattern recognition has been developed. The neural network allows us to find different patterns in the stock market. In addition. several studies were conducted to determine the best parameters for the search and the quality of the patterns found. The algorithm for historical patterns is based on scanning the same price chart with different window sizes to ensure we get all the existing patterns. At the same time, the algorithm for current patterns detection is based on studying a time window that ends on the current day and progressively narrowing it down to find the pattern or the same one but with greater precision. The developed tool also incorporates an alternative pattern search technique based on Dynamic Time Warping, which falls beyond the scope of this paper. The results demonstrate that this technique enables the creation of a functional and valuable tool for investors, allowing them to accurately determine the success rates of various chart patterns over specific time periods. Additionally, the tool facilitates analysis at different levels, whether for individual companies, sector-based groupings, or customized selections based on investor preferences.

 Table 1. Results for Technological and Financial Companies.

	Technological companies		Financial companies	
Chart Patterns	Found	Target	Found	Target
Ascending triangle	110	51 %	118	43 %
Descending triangle	68	44 %	97	42 %
Double bottom	184	62 %	115	53 %
Double top	212	67 %	163	60 %
Head and shoulder	43	65 %	35	69 %
Inverted Head and shoulders	70	69 %	62	69 %

5. Future Work

This research is a work in progress, and several aspects require further exploration. For future work, we aim to expand the range of detected patterns. The tool will be enhanced to display stop-loss levels and target prices, as well as to indicate the most promising and lowest-risk investments using color-coded signals. Another key area for future study is the potential biases in synthetic data generation, particularly analyzing whether artificially modifying patterns affects real-world generalization. Additionally, future work will focus on training the model with greater computational power to enhance its performance and accuracy. These improvements will contribute to a more robust and reliable approach in future studies. Additionally, we plan to conduct studies across different time periods and market sectors to evaluate the effectiveness and reliability of chart patterns over time.

Acknowledgements

Research supported by the framework of the Recovery, Transformation, and Resilience Plan funds, financed by the European Union (Next Generation) under the strategic project C064/23 SCITALA.

References

- [1]. S. B. Achelis, Technical Analysis from A to Z, *McGraw Hill*, 2001.
- [2]. R. T. Farias, J. L. Silva, et al., A literature review of technical analysis on stock markets, *Quarterly Review* of Economics and Finance, Vol. 66, 2017, 2017, pp. 115-126.
- [3]. E. R. Dawson, J. M. Steeley, On the existence of visual technical patterns in the UK stock market, *Journal of Business Finance and Accounting*, Vol. 30, 2003, pp. 263-293.
- [4]. G. C. Friesen, P. A. Weller, et al., Price trends and patterns in technical analysis: A theoretical and empirical examination, *Journal of Banking and Finance*, Vol. 33, 2009, pp. 1089-1100.

(029)

Prediction of Total Daily Diaper Changes Based on Infants' Bowel Sounds During the Beginning of Breastfeeding

<u>S. Mukaiyama</u>¹, N. Tanabe¹ and Y. Oka²

¹ Suwa Tokyo University of Science, Graduate School of Engineering and Management, Nagano, Japan ² Ehime University, School of Medicine, Ehime, Japan Tel.: + 0266 73 1201 E-mail: GH24528@ed.sus.ac.jp, nari@rs.sus.ac.jp

Summary: The management of infant excretion is a vital component of childcare, with significant implications for digestive health and overall well-being. Although traditional methods primarily focus on diagnosing gastrointestinal disorders, the real-time prediction of total daily diaper changes based on bowel sounds remains an underexplored area. To address this gap, we proposed a prediction method utilizing bowel sounds data recorded within the first 10 minutes after morning feeding. Using a custom-designed bowel sound sensor, data were collected from 12 infants (aged 2 to 11 months) over 49 days. The data were analyzed for six distinctive bowel sound features, and the number of sounds within a 10-minute window was counted, reflecting the volume of bowel sounds. By summing the three lowest bowel volume features, we subsequently established ten classification patterns to correlate these features with the total number of daily diaper changes. This approach provides an accurate prediction of diaper change frequency, supporting mothers in planning outings and reducing their caregiving burdens.

Keywords: Infant, Bowel sounds, Diaper change, Prediction, Classification, Breastfeeding, Childcare support.

1. Introduction

Managing infant excretion is an essential aspect of childcare, playing a crucial role in maintaining both the infant's health and overall well-being. The frequency and timing of excretion are closely linked to an infant's gastrointestinal function and overall health status, as highlighted in previous studies [1, 2]. Among the various biological signals reflecting gastrointestinal activity, bowel sounds have gained attention as a non-invasive and real-time indicator of digestive function. Recent research has explored the potential applications of bowel sound analysis in assessing digestive conditions and monitoring excretory behavior. Such analyses have been utilized in the diagnosis of gastrointestinal disorders and in general health management.

In the context of childcare, diaper changes constitute a significant source of stress for caregivers, particularly during outings. Surveys indicate that approximately 80 % of mothers experience concerns regarding diaper changes when outside the home [3]. The unpredictable nature of an infant's excretion schedule makes daily planning difficult and adds to caregivers' physical and mental burden. Furthermore, in Japan, the total fertility rate reached a historic low of 1.20 in 2023 [4], underscoring the urgent need for enhanced childcare support. As societal trends continue to shift toward lower birth rates, technological advancements in childcare support systems are becoming increasingly important for alleviating parental burdens. Despite these challenges, research on predicting infants' daily excretion patterns based on bowel sounds remains limited. While previous studies have explored methods for analyzing

bowel sounds to assess gastrointestinal health, few have investigated their potential in predicting excretion frequency. In particular, the feasibility of predicting long-term excretion trends using short-term bowel sound data has not been thoroughly examined. The extent to which bowel sounds contribute to excretion prediction remains unclear, and there is a lack of research on how bowel sound-based predictive systems could be integrated into childcare support frameworks to reduce caregiver stress. Addressing these research gaps could provide significant benefits in both parenting and broader healthcare applications.

This study aims to develop a predictive model for estimating the total number of diaper changes in a day based on bowel sound data recorded during the first 10 minutes after the morning feeding. This hypothesis is grounded in existing findings that bowel sounds reflect digestive activity and are indicative of subsequent excretion patterns. By extracting and analyzing key features from bowel sound data, we seek to improve the accuracy of excretion frequency prediction. A reliable prediction system would allow caregivers to anticipate diaper changes in advance, facilitating better daily planning and reducing the uncertainty associated with infant excretion.

The findings of this research are expected to contribute to the advancement of data-driven childcare support systems, offering new insights into infant excretion behavior and its relationship with gastrointestinal activity. This study seeks to establish a novel approach to excretion prediction, ultimately providing practical solutions for caregivers and enhancing the overall quality of childcare in modern society.

2. Methods

2.1. Ethics

This study has been approved by the Emergency Research Ethics Committee based on the "Medical Research Involving Human Subjects" guidelines of Ehime University Hospital (Approval No. 1810003).

2.2. Data Collection

In this research, bowel sounds were assessed using a custom-designed device developed under medical guidance, affixed inside an infant's diaper. This device was distributed to each household, and parents were asked to use it for the measurements. The device is affixed by the mother during the measurement process. During feeding sessions, which typically last around 10 minutes, infants remain in a relatively stable state, enabling the observation of characteristic bowel sounds during this period. Additionally, the device collects sounds while minimizing noise caused by the infant's body movements. As a result, bowel sounds were recorded for 10 minutes following each feeding session. Furthermore, the total daily diaper change count (TDDC) was systematically recorded. In collaboration with physicians, six distinct bowel sounds were classified based on their duration and maximum frequency (MF), as outlined in Table 1.

Fable	1.	6	bowel	sounds.
-------	----	---	-------	---------

	MF [kHz]	Duration [ms]
BS1	0.80	0.5
BS2	0.35	2.0
BS3	1.00	-
BS4	0.90	-
BS5	0.30	0.5
BS6	0.40	1.0

2.3. Extraction of Bowel Sounds

For the analysis of bowel sounds recorded during the first 10 minutes of the breakfast session, the Short-Time Fourier Transform (STFT) was applied to the bowel sounds x[n] using Equation (1).

$$X[n,\omega] = \sum_{m=0}^{N} x[n+m]\omega[m]e^{-i\omega m}$$
(1)

In this process, $\omega[m]$ represents the window function, which is a Hamming window, with a sampling frequency of 16,000 Hz, a window size of 256, and a window shift size of 64. For the power spectrum calculation, Equation (2) was utilized.

$$S_{t,h} = 10 \log_{10}(|X[n,\omega]|^2)$$
(2)

This allowed for the visualization of bowel sounds using a spectrogram, as illustrated in Fig. 1.



Fig. 1. Visualisation of observed sounds.

For the state shown in Fig. 1, a significant amount of noise was present, necessitating noise removal. As bowel sounds typically occur at frequencies above 100 Hz [5, 6], low-frequency components below 100 Hz were removed using Equation (3) to mitigate noise interference.

$$h \le 100: S_{t,h} = -80$$
 (3)

Additionally, to minimize environmental noise, power spectra with bowel sounds that greater than environmental noise were obtained. The threshold (5), which allowed for the most accurate extraction of bowel sounds, was used. Therefore, components with spectral intensities below the threshold (5) were excluded using Equation (4).

$$S_{t,h} \le 5: S_{t,h} = -80$$
 (4)

This noise reduction resulted in a spectrogram, as shown in Fig. 2.



Fig. 2. Observed sound after noise reduction.

However, as shown in Fig. 3(a) and Fig. 4(a), this noise reduction process may lead to the loss of some bowel sounds; hence, a compensation procedure was implemented. This compensation procedure includes both "time compensation" and "frequency compensation." Time compensation involves filling in the missing time τ in the power spectrum $S_{t-\tau,h}$ by using the power spectrum immediately before the gap $S_{t,h}$, such that Equation (5). In this process, the allowable time loss is set to 0.048 seconds or less.

7th International Conference on Advances in Signal Processing and Artificial Intelligence (ASPAI' 2025), 8-10 April 2025, Innsbruck, Austria

$$S_{t,h} = S_{t-\tau,h} \tag{5}$$

The results of this compensation are shown in Fig. 3(b).



Fig. 3. Time compensation.

Similarly, frequency compensation involves filling in the missing frequency η in the power spectrum $S_{t-\eta,h}$ by using the power spectrum immediately before the gap $S_{t,h}$, such that Equation (6). In this process, the allowable frequency loss is set to 60 Hz or less.

$$S_{t,h} = S_{t-\eta,h} \tag{6}$$

The results of this compensation are shown in Fig. 4(b).



Fig. 4. Frequency compensation.

The six types of bowel sounds are extracted based on the method outlined in Table 1. The extraction process involves using the maximum frequency, continuous time, and frequency bandwidth for power spectra above a certain threshold. There are three steps in the extraction procedure: (i) checking whether each feature exceeds the maximum frequency, (ii) examining the continuous frequency bandwidth with power spectrum values above a certain threshold, and (iii) examining the duration of continuous time with power spectrum values above a certain threshold. After performing steps (i) to (iii), sounds that meet the conditions for each feature are identified as bowel sounds.

The number of occurrences of each of the six bowel sounds during the first 10 minutes after feeding is then counted and referred to as the "Bowel Sounds Count (BSC)." This methodology allows for the extraction of bowel sound volumes (BSC1 to BSC6) corresponding to the six identified types.

2.4. Calculation of Four Features

The BSC varies among individuals, leading to variability in the data. To account for individual differences, each bowel sound volume was normalized by the total bowel sound volume across all six categories. This standardized measure was termed the Bowel Sounds Proportion (BSP). Since a higher BSP value is considered to have a greater impact on bowel motility, we defined a new feature, BSP7, by summing the BSP values of BS3, BS4, and BS6, which are characterized by lower BSP values. This modification enhances the emphasis on bowel sounds with higher BSP values. In subsequent analyses and discussions, four features (BSP1, BSP2, BSP5, and BSP7) are used.

2.5. 10 Pattern Classifications

Data processing was conducted for 49 days, as outlined in Sections 2.2 and 2.3. The four extracted features were visualized, with the horizontal axis representing BSP and the vertical axis indicating TDDC, as shown in Fig. 5. Notably, the feature patterns exhibited variability even when TDDC values remained constant, indicating the need for classification. Specifically, TDDC values of 7 and 8 were each divided into two groups, while TDDC value 9 was further subdivided into three categories, resulting in ten distinct classification patterns. These patterns were assigned numerical labels (Nos.). Subsequently, interquartile ranges were computed for the four features within each pattern, and the medians were used to generate the graph depicted in Fig. 1. In this figure, the horizontal axis represents the classification numbers and TDDC, while the vertical axis represents BSP. The findings suggest that TDDC can be estimated based on the correlation between the four BSP features and the classification patterns shown in Fig. 6.

3. Simulations

To assess the efficacy of the proposed methodology, bowel sounds were recorded using a compact device during the initial 10 minutes post-feeding. Following the processing of the acoustic data as described in Sections 2.3 and 2.4, the data were grouped into ten distinct classification patterns. The results indicated that the proposed method achieved a prediction accuracy of 83 % in forecasting TDDC.

4. Conclusions

This study aimed to alleviate mothers' childcare responsibilities by establishing a system to predict Total Daily Diaper Changes (TDDC) using infant bowel sounds data. The method developed extracted key features from bowel sound recordings taken during the first 10 minutes after feeding, categorizing them into ten distinct patterns. The results demonstrated high predictive accuracy, indicating that bowel sounds can serve as a reliable indicator for excretion prediction.

These findings suggest that the approach could assist in organizing outings and daily activities, reducing the mental and physical burdens on mothers by providing accurate predictions of diaper changes. This predictive capability helps caregivers plan their day with less uncertainty regarding infant excretion patterns, thereby easing their caregiving responsibilities. Moreover, this study underscores the potential of bowel sound analysis as an effective, noninvasive tool for monitoring gastrointestinal health and predicting excretion behavior. The integration of such data-driven systems in childcare could enhance support for caregivers and contribute to improved infant health management.



Fig. 5. 10-pattern classification.



Fig. 6. Plot 49 days of data.

In conclusion, this research presents a novel method for predicting excretion patterns, which can benefit both caregivers and healthcare systems. Future research could examine broader applications and integrate this method with other health-monitoring technologies to further improve childcare and healthcare support.

5. Limitations and Future Directions

The study has five limitations.

In this study, we developed a system capable of predicting the total number of diaper changes per day. However, the system does not estimate the specific timing of diaper changes, which may not provide sufficient information for mothers. Therefore, a future direction of this research is to construct a system that predicts the timing of infant bowel movements. There is variability in the number of diaper changes due to the inclusion of infants ranging from one to eleven months old, as well as differences in gender and dietary intake. To address this, future work aims to develop personalized predictive models tailored to individual infants.

This study only uses data from 12 infants over a span of 49 days, which limits the statistical reliability of the findings. Therefore, to evaluate the generalizability of the model, larger datasets will be required for future validation.

The goal of this research was to predict the total number of diaper changes per day. In the future, we plan to develop a mobile application that notifies mothers of the prediction results in real-time, which would help reduce their burden.

Challenges such as environmental noise and infant movements were also identified. To improve accuracy, better noise reduction techniques and more sensitive sensors will be necessary.

References

- G. J. Jordan, Elimination communication as colic therapy, *Medical Hypotheses*, Vol. 83, Issue 3, 2014, pp. 282-285.
- [2]. B. D. H. Brandon, D. Hatch, A. Barnes, A. J. Vance, J. Harney, B. Voigtman, N. Younge, Impact of diaper change frequency on preterm infants' vital sign stability and skin health: A RCT, *Early Human Development*, Vol. 164, 2022, 105510.
- [3]. Unicharm, Everyone's Concerns about Changing Baby Nappies on the Go, MOONY, https://jp.moony.com/ja/ tips/baby/childcare/diapers/bt0836.html

7th International Conference on Advances in Signal Processing and Artificial Intelligence (ASPAI' 2025), 8-10 April 2025, Innsbruck, Austria

- [4]. 2023 Overview of the Vital Statistics Monthly Report, Total Fertility Rate, *Ministry of Health, Labour and Welfare*, 2024 (in Japanese).
- [5]. T. Yoshino, H. Yoshino, H. Kanno, T. Kusaba, Spectral analysis of bowel sounds in ileus, *Journal of Japan Surgical Association*, Vol. 44, Issue 12, 1985, pp. 1583-1592.
- [6]. S. S. Ching, Y. K. Tan, Spectral analysis of bowel sounds in intestinal obstruction using an electronic stethoscope, *World Journal of Gastroenterology*, Vol. 18, Issue 33, 2012, pp. 4585-4592

(030)

Deep Jansen-Rit Parameter Inference for Model-driven Analysis of Brain Activity

Deepa Tilwani, and Chrisitan O'Reilly

Department of Computer Science and Engineering, University of South Carolina, Columbia, SC, USA Artificial Intelligence Institute, University of South Carolina, Columbia, SC, USA Carolina Autism and Neurodevelopment Research Center, University of South Carolina, Columbia, SC, USA Institute for Mind and Brain, University of South Carolina, Columbia, SC, USA E-mail: dtilwani@mailbox.sc.edu, christian.oreilly@sc.edu

Summary: Accurately modeling effective connectivity (EC) is critical for understanding how the brain processes and integrates sensory information. Yet, it remains a formidable challenge due to complex neural dynamics and noisy measurements such as those obtained from the electroencephalogram (EEG). Model-driven EC infers local (within a brain region) and global (between brain regions) EC parameters by fitting a generative model of neural activity onto experimental data. This approach offers a promising route for various applications, including investigating neurodevelopmental disorders. However, current approaches fail to scale to whole-brain analyses and are highly noise-sensitive. In this work, we employ three deep-learning architectures—a transformer, a long short-term memory (LSTM) network, and a convolutional neural network and bidirectional LSTM (CNN-BiLSTM) network—for inverse modeling and compare their performance with simulation-based inference in estimating the Jansen-Rit neural mass model (JR-NMM) parameters from simulated EEG data under various noise conditions. We demonstrate a reliable estimation of key local parameters, such as synaptic gains and time constants. However, other parameters like local JR-NMM connectivity cannot be evaluated reliably from evoked-related potentials (ERP). We also conduct a sensitivity analysis to characterize the influence of JR-NMM parameters on ERP and evaluate their learnability. Our results show the feasibility of deep-learning approaches to estimate the subset of learnable JR-NMM parameters.

Keywords: EEG, Neural mass model, LSTM, Transformers, Simulation-based inference, Model-Driven analysis.

1. Introduction

Neural mass models (NMMs) have emerged as powerful computational tools for simulating collective neuronal behavior at a mesoscopic scale, offering a mathematically tractable approach to modeling brain dynamics for the past five decades [12, 26]. These models provide a crucial bridge between microscopic neural activity and macroscopic neuroimaging signals, enabling quantitative analysis of normal brain functions and pathophysiological conditions [5]. NMMs create computationally efficient abstractions that can be simulated at scale across brain regions by capturing the dynamics of the average activity of neuronal populations as systems of coupled differential equations. Integrating NMMs with computational data analysis pipelines has enabled significant advances in neuroimaging [19, 4]. Modern implementations of these models allow researchers to simulate the transition between stable and pathological brain states, providing a basis for biomarker development and quantitative analysis of neurological disorders. NMMs can be particularly effective for modeling epileptiform activity [25], oscillatory disturbances [11], and state transitions in disorders like Alzheimer's disease [21].

Effective connectivity (EC) captures the causal influence that neural systems exert on one another. Unlike functional connectivity, which merely describes statistical dependencies, EC captures the directional influence of one neural population on another. Dynamic causal modeling (DCM) [4] represents a principled algorithmic approach to estimating these causal relationships from neural signals. This approach frames the problem as a Bayesian model inversion and explains observed data through a generative model of coupled neural masses. Bayesian techniques and, hence, DCM [7, 3, 6] are computationally demanding and, therefore, are generally inadequate for handling large datasets or real-time analysis.

Simulation-based inference (SBI) has recently emerged as a computationally efficient approach for parameter estimation in complex dynamical systems [22]. SBI algorithms, including Sequential Neural Posterior Estimation (SNPE) [15], leverage advances in amortized inference to approximate posterior distributions without requiring explicit likelihood calculations. These methods can handle complex forward models but face diminishing performance as the parameter spaces increase, creating computational bottlenecks for whole-brain analyses.

Deep learning architectures present a promising paradigm for parameter recovery in complex dynamical systems. Recent advances in neural network architectures - particularly sequence models like Transformers [1] and Long Short-Term Memory (LSTM) [27] networks can efficiently learn the mapping between model parameters and observable outputs. These networks compress the high-dimensional functional relationship between parameters and outputs, enabling rapid inference once trained. Recent research has successfully applied these techniques to connectome-based NMMs for functional magnetic resonance imaging (fMRI) [8], suggesting broader applicability to neuroimaging data analysis.

In this paper, we systematically evaluate four approaches for parameter inference in the Jansen-Rit Model (JR-NMM): (1) Neural Mass an EEGTransformer architecture using multi-head attention mechanisms and Transformer's Encoder components only, (2) an LSTM network specialized for sequential data processing, (3) a convolutional neural network and bidirectional LSTM (CNN-BiLSTM), and (4) the SBI approach using SNPE. We benchmark these algorithms' performance in estimating nine JR-NMM parameters (as defined later, in Table 1) from EEG data under varying noise conditions. Through comprehensive computational experiments and sensitivity analyses, we establish the relative strengths of each approach and identify which parameters can be recovered from EEG signals reliably. This computational benchmarking study lays essential groundwork for developing robust algorithms that infer local and global parameters from empirical EEG recordings.

2. Background

Brain dynamics span multiple spatial and temporal scales. Accordingly, its computational modeling ranges from detailed biophysical simulations of individual neurons to abstract population-level models of regional activity [4, 16, 17]. NMMs occupy a crucial middle ground in this landscape, providing computationally tractable abstractions of neural population dynamics anchored in neurophysiological mechanisms. These models distill the essential properties of neural circuits while remaining amenable to efficient numerical simulation, making them valuable tools for neuroinformatics and computational neuroscience.

When formalized in 1995 [10], JR-NMM represented a significant advance in mesoscopic brain modeling. This model extended earlier mathematical formulations by Lopes da Silva [12] by implementing a cortical column as a system of coupled differential equations representing interacting populations of pyramidal cells, excitatory interneurons, and inhibitory interneurons. This architecture enables the simulation of various EEG rhythms and event-related potentials (ERPs) through appropriate parameter configurations. computational efficiency and biological The plausibility of models like JR-NMM made them foundational components in whole-brain simulation platforms, including frameworks like The Virtual Brain (TVB) [18], neurolib [2], and FastDMF [19].

The JR-NMM is formulated as a system of three second-order ordinary differential equations (ODEs) and generally transformed into six first-order ODEs.

This formulation enables efficient numerical integration using standard ODE solvers such as Euler or Runge-Kutta methods. The model parameters – which include synaptic gains (A_e , A_i), time constants (b_e , b_i), connectivity strengths (a_1 - a_4), and an overall scaling factor (C) - determine the dynamics of the system. Variations in these parameters can produce diverse computational behaviors, including fixed points, limit cycles, and chaotic attractors, corresponding to different patterns observed in the electroencephalogram (EEG) [20].

Challenges in parameter estimation in NMMs stem from several algorithmic and mathematical factors. First, the nonlinear nature of these models creates a complex relationship between parameters and observable outputs. Second, the models exhibit partial identifiability, where multiple parameter combinations can produce nearly identical outputs, making their fitting a mathematically ill-posed inverse problem. Third, the high-dimensional parameter space creates a computationally intensive search problem that scales poorly with traditional optimization methods. Finally, the stochastic nature of neural recordings due to measurement noise and unmeasured inputs introduces additional complexity.

Conventional approaches to NMM parameter estimation have primarily relied on Bayesian inference. DCM [4] implements a variational Bayesian algorithm to approximate posterior distributions over model parameters. While mathematically elegant, this implementation relies on gradient-based optimization under a linear approximation of model dynamics, limiting its applicability to large-scale models or highly nonlinear systems [17]. The computational complexity of this algorithm increases exponentially with the number of brain regions, making whole-brain computationally prohibitive analyses without substantial high-performance computing resources.

Recent advances in SBI [22] have introduced new paradigms for parameter estimation in complex dynamical systems. Unlike traditional Bayesian methods that require explicit likelihood functions, SBI algorithms like SNPE approximate posterior distributions directly through repeated simulations [15]. These methods leverage neural density estimators to learn the conditional distribution of parameters given observed data. While SBI provides greater flexibility for complex forward models, the rigid requirements for the experimental data to closely match the generated simulation data in all aspects (i.e., noise characteristics, free parameters, exact priors, and other underlying properties) limits this approach [24].

Deep learning architectures offer alternative strategies for parameter recovery in NMMs. Transformer models built on self-attention mechanisms can capture long-range dependencies in time series without the sequential constraints of recurrent networks. Meanwhile, LSTM provides specialized computational structures for processing sequential information through gated memory cells that can retain information over an extended period. Both architectures can effectively learn complex mappings between model parameters and observable outputs when trained on sufficiently large simulation datasets.

The study of effective connectivity (EC) extends beyond methodological considerations, addressing fundamental questions about information processing in neural systems. EC quantifies the causal influence neural populations exert on each other, providing insights into functional integration and segregation within the brain. Through analyses of neuroimaging data, disruptions in EC have been implicated in various neurological and psychiatric conditions. Developing computationally efficient and robust methods for EC inference could enhance both the theoretical understanding of brain organization and practical applications in clinical neuroscience. Our work extends the computational neuroscience literature by systematically benchmarking four distinct algorithmic approaches to local parameter inference in the JR-NMM. We establish each method's relative computational efficiency and accuracy by comparing EEGTransformer, LSTM, and CNN-BiLSTM architectures, and an SNPE-based simulation approach across different noise conditions. This comparative analysis provides crucial insights into which parameters are most reliably recoverable from EEG data. The resulting benchmark offers a foundation for developing more sophisticated algorithms to infer local (within a brain region) and global (between brain regions) parameters in whole-brain simulations based on NMMs.

3. Methods

Neural Mass Model: We implemented the JR-NMM [10] to simulate cortical activity. It is governed by the system of coupled differential equations (1). In this system, indices 0, 2, and 4 represent the pyramidal, excitatory, and inhibitory neuron populations, respectively. The output signal $y(t) = a_2y_2 - a_4y_4$ represents the difference between the pyramidal's excitatory and inhibitory postsynaptic potentials. This value is a proxy for EEG sources because EEG is thought to reflect mainly the postsynaptic potentials in the apical dendrites of pyramidal cells [14]. Table 1 shows the default parameter values and ranges used in our simulations based on physiologically plausible values from previous studies [10, 4].

Simulation Protocol: The simulations were performed with a temporal resolution of 1 ms. Each simulation consisted of 1) an 800 ms transient period to allow the system to reach a steady state, 2) a 200 ms baseline period before stimulus onset, and 3) stimulus events with an inter-stimulus interval of 1 second, presented 60 times. We modeled the stimulus with a

50 ms wide rectangular pulse with a 60 mV amplitude for pyramidal cells and a 30 mV amplitude for inhibitory interneurons. Moran et al. [13], we incorporated voltage-dependent synaptic mechanisms in inhibitory interneurons to better capture feedforward inhibition dynamics. The primary differential equations were solved using a forward Euler integration method.

Table 1. Standard parameter settings for Jansen-Rit model.
Only the value from parameters with a range is estimated
in this study.

Parameter	Description	Default	Range
Ae	Excitatory gain	3.25 mV	2.6-9.75 mV
Ai	Inhibitory gain	22 mV	17.6-110 mV
be	Excit. time const.	100 s ⁻¹	5-150 s ^{- 1}
bi	Inhib. time const.	50 s ⁻¹	25-75 s ^{- 1}
С	Connect. const.	135	65-1350
a 1	Connect. param.	1.0	0.5-1.5
a ₂	Connect. param.	0.8	0.4-1.2
a ₃	Connect. param.	0.25	0.125-0.375
a 4	Connect. param.	0.25	0.125-0.375
Vmax	Max firing rate	5 s ^{- 1}	—
v	Firing threshold	6 mV	_
r	Sigmoid steepness	0.56	_

From Neural Activity to EEG: We implemented an EEG forward model to transform NMM output into realistic EEG signals. The output signal from the JR-NMM (difference between pyramidal and inhibitory potentials) was scaled by a gain factor of 10^{-6} to account for the scale simulated signals to physiologically realistic EEG amplitudes (10-100 μ V). This signal was then used as a source located to a specific cortical region ("caudalmiddlefrontal-lh") using the FreeSurfer average (fsaverage) brain template, as defined in MNE-Python. A forward solution was computed using the Boundary Element Method (BEM) to model volume conduction. Sensor projections were created for a standard 64-channel BioSemi EEG montage. To test the robustness of our inference algorithms under different signal-to-noise ratio (SNR) conditions, we added noise to the simulated EEG signals using a scaling approach. We scaled the *ad hoc* noise covariance matrix, generated using MNE's *make ad hoc cov function*, with a factor varying from 0 to 1 in steps of 0.1 to simulate different noise levels added to synthetic EEG signals. A noise factor of 0 represented noise-free simulations, while a factor of 1.0 applied the full noise covariance, resulting in realistic noise levels comparable to empirical recordings. ERPs were extracted by creating epochs around each stimulus event (from -200 ms to 1000 ms), applying baseline correction using the pre stimulus interval (-200 to 0 ms), and averaging all 60 trials.

Parameter Inference Models: We implemented four distinct approaches for inferring the JR-NMM parameters from EEG data, as summarized in Fig. 1. We developed the EEGTransformer model based on Transformer's Encoder components only, which processes EEG signals through an attention-based architecture with eight attention heads, an embedding dimension of 256, and a multi-layer feed-forward network (1024 \rightarrow 256) with a dropout regularization (0.2). The Vanilla LSTM model treats EEG data as sequential information using a 2-layer network with 64 hidden units and dropout regularization (0.2). The CNN-BiLSTM integrates a CNN for feature extraction with bidirectional LSTMs for temporal dependency analysis. It employs a hierarchical 1D CNN structure (filters: $32 \rightarrow 64 \rightarrow 128$, kernels: $7 \rightarrow 5 \rightarrow 3$) with max pooling, followed by bidirectional and unidirectional LSTM layers (128 units each), multiple dropout layers (0.3), and dense layers (128 \rightarrow 64) culminating in a linear output layer. This architecture was designed to capture both spatial patterns across channels and complex temporal dynamics in ERP signals. All neural network models were trained for 50 epochs with batch size 32. The SBI approach uses a BoxUniform prior and sequential neural posterior estimation. To train, we generated a dataset of 1000 simulations per noise level. All four methods were used to estimate the same 9 JR-NMM parameters (Ae, Ai, be, bi, a1-a4, C).



Fig. 1. Flow diagram for the inference of A_e , A_i , b_e , b_i , a_1 - a_4 , and C using the EEGTransformer, LSTM, CNN-BiLSTM, and SBI methods. The JR model generates simulated EEG for training and testing.

Data Processing Pipeline: We adopted a supervised regression approach for parameter inference, following standard deep learning practices [3]. Table 2 outlines our processing pipeline, designed to optimize neural network training for high-dimensional EEG data.

Implementation Details: Models were implemented in Python 3.12 with PyTorch 1.9. Jansen-Rit neural simulations were mapped to scalp EEG through an EEG forward model implemented using MNE-Python 1.9.0. Additional libraries included xarray for data storage, tqdm for progress tracking, seaborn for visualization, and scikit-learn for preprocessing and evaluation.

Table 2. Data processing steps.

Stage	Details
Database	1000 simulations, 64-channel EEG
Creation	60 trials per simulation
Normalization	JR-NMM parameters normalized to [0,1]
Data Split	80 % training, 10 % validation, 10 % testing (random state = 68)
Evaluation	Pearson correlation

Sensitivity Analysis: We conducted a sensitivity analysis to quantify how variations in Jansen-Rit parameters affect simulated EEG. For each parameter $(A_e, A_i, b_e, b_i, a_1-a_4, C)$, we performed 200 simulations while varying the target parameter from the range mentioned in Table 1, keeping all other parameters fixed. We generated ERPs for each parameter configuration without noise. This approach allowed us to isolate each parameter's unique influence on EEG output. For quantitative assessment, we quantified parameter sensitivity through three metrics:

- Raw evoked potentials: $\Delta ERP(p_i, t)$, where p_i is the parameter, t is time;
- Deviations from mean: $\Delta \text{ERP}(p_i, t) = \text{ERP}(p_i, t) \frac{1}{N} \sum_{j=1}^{N} \text{ERP}(p_j, t)$, with N = 200 parameter values;
- Gradient with respect to parameter: $\nabla_{p} ERP(p_{i}, t) = \frac{\partial ERP(p, t)}{\partial p} |$.

These metrics were visualized as heatmaps showing their variation depending on parameter values and time.

4. Results

4.1. Results of Sensitivity Analysis

Sensitivity analysis revealed marked differences in parameter influence on simulated ERPs (Fig. 2 and Fig. 3). These sensitivity heatmaps highlight how each parameter shapes the temporal profile of the ERP. The raw ERP demonstrated that excitatory parameters (Ae, a_1, a_2) primarily increased the peak amplitude around 250ms, while inhibitory parameters (A_i, a₃, a₄) have an opposite effect. The error (deviation from the mean response) exhibited clear parameter-specific temporal windows of influence: Ae and Ai showed unimodal patterns whereas be and bi showed an alternance of positive and negative peaks. The gradient further quantified sensitivity magnitude, revealing that Ai had a more significant influence (about an order of magnitude) than Ae on ERP morphology, despite both being gain parameters. Key takeaways include: (1) connectivity parameters exhibit functional differentiation, with a₁, a₂ and a₃, a₄ showing opposite gradient polarities; (2) the global coupling parameter C affects multiple aspects of ERP morphology with complex biphasic patterns; and (3) excitatory time constant b_e demonstrated distributed temporal effects

(i.e., longer tail), indicating its crucial role in overall signal timing. These findings provide essential insights for parameter estimation and model interpretation in neurophysiological studies.



Fig. 2. Sensitivity analysis for synaptic parameters (A_e, A_i, b_e, b_i) showing the ERP amplitude (left), error (middle), and gradient (right).



Fig. 3. Sensitivity analysis for local connectivity (a1-a4) and C showing the ERP amplitude (left), error (middle), and gradient (right).

4.2. Parameter Inference Performance

The parameter inference capabilities of our four models - EEGTransformer, CNN-BiLSTM, Vanilla LSTM, and SBI - were evaluated across varying noise conditions (Fig. 4). Results show distinct patterns of parameter recoverability across the different architectures. The inhibitory parameter bi demonstrated the strongest recovery performance across multiple models. The EEGTransformer and CNN-BiLSTM maintained exceptionally high correlation coefficients (>0.9) for this parameter even at high noise levels, with SBI showing similarly strong performance. In contrast, the Vanilla LSTM performed poorly on this parameter, with correlations fluctuating around 0.1-0.2, suggesting fundamental limitations in capturing temporal dependencies (i.e., adding a 1D CNN to meaningful temporal features allowed the CNN-BiLSTM to perform much better than the vanilla LSTM). The excitatory parameter b_e showed moderate recoverability, with the EEGTransformer achieving consistent correlations of approximately 0.5 across all noise conditions. The CNN-BiLSTM and SBI showed more variable performance for this parameter, with correlations typically in the 0.1-0.3 range, while Vanilla LSTM exhibited inconsistent correlations between 0.1-0.5. The connectivity parameters (a_1-a_4) exhibited poor recoverability across all models, with correlations rarely exceeding 0.2 and often oscillating around zero. Parameter C showed better recoverability with the EEGTransformer than other models, maintaining positive correlations around 0.2 across noise levels. The EEGTransformer and CNN-BiLSTM demonstrated superior robustness to noise compared to both the Vanilla LSTM and SBI approaches.



Fig. 4. Comparison of deep learning models estimating JR-NMM parameters across noise levels. EEGTransformer and CNN-BiLSTM maintaining high correlations for b_i. The vanilla LSTM underperforms with lower correlations, while SBI exhibits parameter dependent variability.

5. Discussion

Our comprehensive benchmarking study offers several key insights into the challenging problem of NMMs parameter inference from EEG data. NMM parameter estimation allows us to study the underlying neurophysiological mechanisms generating EEG. By estimating JR-NMM parameters from EEG, we directly map measurable brain signals and the biophysical properties of neural populations, enabling a mechanistic understanding of cortical dynamics and the potential identification of aberrant parameter configurations in pathological conditions. The sensitivity analysis revealed a clear hierarchy of parameter identifiability within the JR-NMM, with synaptic gains and time constants exhibiting stronger ERP signatures than connectivity parameters. This finding aligns with the fundamental architecture of the JR-NMM, where synaptic parameters directly modulate signal amplification and decay, creating more distinctive output patterns. The sensitivity analysis also clarifies why some parameters show poor identifiably. For example, A_i and A_e have almost identical patterns, only with inverse amplitude. Therefore, it is difficult to determine whether a reduction in ERP amplitude is caused by a decrease in Ae or an increase in Ai. Consequently, in terms of identifiability, the JR model may provide better results for inverse modeling if we reparametrize it as a function of the excitatory-inhibitory ratio $r_{ei} = A_e/A_i$ and a global offset parameter $\alpha = (A_e + A_i)/2$ such that $A_e = 2\alpha r_{ei}/(1 + r_{ei})$ and $A_i = 2\alpha/(1 + r_{ei})$. Moreover, such a parameterization would map well the report of an imbalance in with the excitatory/inhibitory ratio in various conditions, including autism spectrum disorder [23].

The comparative analysis of inference approaches demonstrated that Transformer-based architectures achieve superior robustness to noise compared to both the LSTM and SBI approaches. This advantage likely stems from the Transformer's self-attention mechanism, which can capture non-sequential dependencies across the entire ERP signal, unlike the strictly sequential processing of LSTMs. The consistent performance of the EEGTransformer, particularly for A_e and b_i parameters, suggests that attention-based architectures may be better suited for extracting relevant features from complex temporal signals like EEG.

The parameter recovery performance broadly corresponds with our sensitivity analysis findings, with parameters showing minimal impact on ERP morphology (a₁-a₄) being poorly recovered by all models, regardless of noise level. This observation confirms the fundamental challenge of parameter identifiability in complex dynamical systems. If parameters do not create distinct signatures in the observable output, no inference method can reliably recover them without additional constraints or information. The SBI approach performed well for some parameters but is highly sensitive to noise for others, suggesting that SBI may be valuable in low-noise conditions or combined with other ensemble-method approaches. The LSTM model's inconsistent performance across noise levels indicates potential limitations in applying recurrent architectures to this problem without additional regularization or architectural modifications.

We must acknowledge several limitations to this study. First, it used simulated data that incorporated realistic noise but could not capture the full complexity of real EEG recordings. Furthermore, we simulated neural activity using a JR-NMM in a single cortical region, while brain activity typically involves distributed networks of interacting regions. The parameter identifiability patterns observed may differ in multi-region models or when using alternative neural mass formulations with different state variables and connectivity architectures.

6. Conclusion

This study demonstrates that Transformer-based and CNN-BiLSTM architectures outperform LSTM and SBI approaches for JR-NMM parameter inference, with synaptic parameters showing higher recoverability than connectivity parameters. Our sensitivity analysis establishes a clear relationship between parameter influence on ERP morphology and inference reliability, providing a foundation for more robust approaches for NMMs parameter estimation. These insights have important implications for computational neuroscience and clinical applications. Future work should extend this analysis to empirical EEG data, explore transfer learning approaches to bridge synthetic and real data domains, and investigate the potential of hybrid architectures that combine the strengths of different inference approaches. Expanding this benchmarking to other neural mass models and exploring multi-modal data integration could enhance parameter identifiability. Further, reparametrizing existing NMM may offer equivalent but more identifiable models. Robust parameter inference methods will ultimately enable more reliable neural dynamics characterization in healthy and pathological states, advancing our understanding of brain function and dysfunction.

Acknowledgements

This work was supported by COR's startup package at USC and an NSF grant to COR (#2419634).

References

- A. Anwar, Y. Khalifa, J.L. Coyle, E. Sejdic, Transformers in biosignal analysis: A review, *Information Fusion*, Vol. 114, 2025, 102697.
- [2]. C. Cakan, N. Jajcay, K. Obermayer, neurolib: A simulation framework for whole-brain neural mass modeling, *Cognitive Computation*, Vol. 15, 2023, pp. 1132-1152

- [3]. K. Cranmer, J. Brehmer, G. Louppe, The frontier of simulation-based inference, *Proceedings of the National Academy of Sciences*, Vol. 117, Issue 48, 2020, pp. 30055-30062.
- [4]. O. David, K. Friston, A neural mass model for MEG/EEG: coupling and neuronal dynamics, *NeuroImage*, Vol. 20, 2003, pp. 1743-1755.
- [5]. G. Deco, V. K. Jirsa, P. A. Robinson, M. Breakspear, K. Friston, The dynamic brain: from spiking neurons to neural masses and cortical fields, *PLoS Computational Biology*, Vol. 4, Issue 8, 2008, e1000092.
- [6]. A. Gelman, J. Hill, Data Analysis Using Regression and Multilevel/Hierarchical Models, *Cambridge University Press*, 2006.
- [7]. D. Greenberg, M. Nonnenmacher, J. Macke, Automatic posterior transformation for likelihood-free inference, in *Proceedings of the International Conference on Machine Learning*, 2019, pp. 2404-2414.
- [8]. J. D. Griffiths, Z. Wang, S. H. Ather, D. Momi, S. Rich, A. Diaconescu, A. R. McIntosh, K. Shen, Deep learning-based parameter estimation for neurophysiological models of neuroimaging data, *bioRxiv*, 2022, 2022-05.
- [9]. R. Herzog, P. A. M. Mediano, F. E. Rosas, A. I. Luppi, Y. Sanz-Perl, E. Tagliazucchi, M. L. Kringelbach, R. Cofré, G. Deco, Neural mass modeling for the masses: Democratizing access to wholebrain biophysical modeling with FastDMF, Network *Neuroscience*, Vol. 8, Issue 4, 2024, pp. 1590-1612.
- [10]. B. H. Jansen, V. G. Rit, Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns, Biological Cybernetics, Vol. 73, 1995, pp. 357-366.
- [11]. L. Kuhlmann, D. R. Freestone, J. H. Manton, B. Heyse, H. E. M. Vereecke, T. Lipping, M. M. R. F. Struys, D. T. J. Liley, Neural mass model-based tracking of anesthetic brain states, *NeuroImage*, Vol. 133, 2016, pp. 438-456.
- [12]. F. H. Lopes da Silva, A. Hoeks, H. Smits, L. H. Zetterberg, Model of brain rhythmic activity: the alpha-rhythm of the thalamus, *Kybernetik*, Vol. 15, 1974, pp. 27-37.
- [13]. R. Moran, D. A. Pinotsis, K. Friston, Neural masses and fields in dynamic causal modeling. *Frontiers in Computational Neuroscience*, Vol. 7, 2013, 57.
- [14]. P. L. Nunez, R. Srinivasan, Electric Fields of the Brain: the Neurophysics of EEG, *Oxford University Press*, 2006.
- [15]. G. Papamakarios, I. Murray, Fast ε-free inference of simulation models with Bayesian conditional density estimation, in *Proceedings of the Conference on Neural Information Processing Systems (NIPS'16)*, 2016, pp. 1-9.
- [16]. D. Pinotsis, P. Robinson, P. Beim Graben, K. Friston, Neural masses and fields: Modelling the dynamics of brain activity, *Frontiers in Computational Neuroscience*, Vol. 8, 2014.
- [17]. S. Sadeghi, D. Mier, M. F. Gerchen, S. N. L. Schmidt, J. Hass, Dynamic causal modeling for fMRI with Wilson-Cowan-based neuronal equations, *Frontiers in Neuroscience*, Vol. 14, 2020, 593867.
- [18]. P. Sanz Leon, S. A. Knock, M. M. Woodman, L. Domide, J. Mersmann, A. R. McIntosh, V. Jirsa, The virtual brain: a simulator of primate brain network

7th International Conference on Advances in Signal Processing and Artificial Intelligence (ASPAI' 2025), 8-10 April 2025, Innsbruck, Austria

dynamics, Frontiers in Neuroinformatics, Vol. 7, 2013, 10.

- [19]. R. C. Sotero, N. J. Trujillo-Barreto, Y. Iturria-Medina, F. Carbonell, J. C. Jimenez, Realistically coupled neural mass models can generate EEG rhythms, *Neural Computation*, Vol. 19, Issue 2, 2007, pp. 478-512.
- [20]. A. Spiegler, T. R. Knösche, K. Schwab, J. Haueisen, F. M. Atay, Modeling brain resonance phenomena using a neural mass model, *PLoS Computational Biology*, Vol. 7, Issue 12, 2011, e1002298.
- [21]. L. Stefanovski, J. M. Meier, R. K. Pai, P. Triebkorn, T. Lett, L. Martin, K. Bülau, M. Hofmann-Apitius, A. Solodkin, A. R. McIntosh, et al., Bridging scales in Alzheimer's disease: Biological framework for brain simulation with the virtual brain, *Frontiers in Neuroinformatics*, Vol. 15, 2021, 630172.
- [22]. A. Tejero-Cantero, J. Boelts, M. Deistler, J.-M. Lueckmann, C. Durkan, P. J. Gonçalves, D. S. Greenberg, J. H. Macke, SBI: A toolkit for simulation-based inference, *Journal of Open Source Software*, Vol. 5, Issue 2, 2020, 2505.

- [23]. G. Uzunova, S. Pallanti, E. Hollander, Excitatory/inhibitory imbalance in autism spectrum disorders: Implications for interventions, therapeutics, *The World Journal of Biological Psychiatry*, Vol. 17, Issue 3, 2016, pp. 174-186.
- [24]. B. Wang, J. Leja, V. A Villar, J. S. Speagle, SBI++: Flexible, ultrafast likelihood-free inference customized for astronomical applications, *The Astrophysical Journal Letters*, Vol. 952, Issue 1, 2023, L10.
- [25]. F. Wendling, P. Benquet, F. Bartolomei, V. Jirsa, Computational models of epileptiform activity, *Journal of Neuroscience Methods*, Vol. 260, 2016, pp. 233-251.
- [26]. H. R. Wilson, J. D. Cowan, Excitatory, inhibitory interactions in localized populations of model neurons, *Biophysical Journal*, Vol. 12, 1972, pp. 1-24.
- [27]. Y. Yu, X. Si, C. Hu, J. Zhang, A review of recurrent neural networks: LSTM cells, network architectures, *Neural Computation*, Vol. 31, Issue 7, 2019, pp. 1235-1270.

(031)

Computing the Time-dependent Krankheit-operator in Epilepsy from ECoG: a Case Study

<u>M. Mannone</u>^{1,2,3,4}, P. Ribino¹, A. Saibene^{5,8}, P. Fazio^{4,6}, S. Fazio⁷, F. Gasparini^{5,8}, M. Gherardi⁷ and N. Marwan^{2,3}

¹ ICAR, National Research Council of Italy (CNR), Italy
 ² Institute of Physics and Astronomy, University of Potsdam, Germany
 ³ Potsdam Institute for Climate Impact Research (PIK), Member of the Leibniz Association, Germany
 ⁴ DSMN, Ca' Foscari University of Venice, Italy
 ⁵ Università degli Studi di Milano-Bicocca, Italy
 ⁶ VSB – Technical University of Ostrava, Czechia
 ⁷ Università Statale di Milano, Italy
 ⁸ NeuroMI, Milan Center for Neuroscience, Milano, Italy
 E-mail: maria.mannone@icar.cnr.it

Summary: Description and prediction of epileptic seizures present numerous challenges that can be addressed with physics and computer science. Here, we investigate epilepsy by developing a specific form of the Krankheit-operator (*K*-operator), a physics-based approach modeling disease-driven damage on brain pathways. The *K*-operator acts on the different layers of the brain, from neurons, to neural agglomerates, to lobes. Its first experimental applications to functional magnetic resonance images dealt with interactions between regions of interest of the brain. Here, we consider the action of *K* between different brain areas described by channels in electrocorticography (ECoG) for the first time. In particular, we focus on temporal lobe epilepsy, applying the methodology to a case study, i.e., the data acquired on a person monitored via pre-surgery ECoG. The information before, immediately before, during, and after an epileptic seizure is encoded in matrices, and investigated with the tools of operatorial algebra adopted in physics, shaping a form of the *K*-operator for our case study. We discuss our preliminary results and sketch further lines of development.

Keywords: Electrocorticography (ECoG), Epilepsy, Krankheit-operator (K-operator), Disease evolution, Seizure detection

1. Introduction

Taken by surprise, maybe by a demon, an external entity: this is why ancient Greeks used the verb ἐπιλαμβάνω, take hold of, or attack, while referring to seizures [1]. And this is the origin of the word epilepsy². Epileptic seizures, which are disabling and the potentially life-threatening for risk of self-suffocation, are the object of studies between neurology, computation, and mathematical modeling [2-7]. Epilepsy is related to the brain's electrical activity and can be measured bv electroencephalograms (EEGs) [8]. Pre-surgical interventions usually rely on intracranial EEGs (iEEGs) and thus require surgeons to place electrodes directly in the brain. This allows a finer monitoring of brain activities at the cost of being invasive. Concerning iEEG, research on epileptic seizure prediction has recently been categorized into four main categories [9]. The first category is related to approaches performing time-domain analysis, which is a linear methodology for directly extracting meaningful information from EEG signals within the temporal domain. This approach is intuitive, has a clear physical interpretation, and is easy to comprehend [10, 11]. The second type of approach concerns the frequency-domain analysis, which facilitates the examination of the neural signal spectral distribution patterns and the differences among frequency components [12]. The third one concerns the time-frequency analysis, integrating both temporal and spectral components. This kind of approach effectively captures the transient information inherent in EEG signals by extracting multi-component features that exhibit time-frequency variations [13, 14]. Finally, most recent approaches rely on brain networks. It is worth noting that the brain can be conceptualized as a sophisticated network in which various regions engage in communication and collaboration to execute diverse functions. It has been demonstrated that abnormal dynamic alterations in brain functional connectivity manifest in certain patients during seizure episodes. Furthermore, notable variations in brain functional connectivity patterns are evident across distinct seizure stages. Brain network analysis serves as a more effective method for elucidating the global characteristics of neural activity in epilepsy, particularly in relation to seizure onset, by examining

² The verb *epilambàno*, infinite *epilambànein*, with the preposition *epi*, was also used to indicate the motion of fruits falling from a tree, without any "will." Similarly, the epileptic crisis is something attacking, involving the patient

in a series of fast and completely involuntary movements. The root $\lambda \alpha \psi$ then became $\lambda \eta \psi$, *leps*, leading to the English word epilepsy.

the degree of synchronization among various brain regions [15, 16]. The approach we propose in this paper aligns with the latter category.

Among the different invasive technologies, one of the most diffused is electrocorticography (ECoG), which is often used to identify the precise source of seizures before proceeding with surgical resections of specific connections [17]. The analysis of ECoG helps improve predictive methodologies holding the potential to mitigate the adverse effects of inherent uncertainties of seizures' occurrences. From single patients to cohort investigations, the prediction of epileptic seizures constitutes a considerable challenge within the domain of neurology. Consequently, there is a growing emphasis among researchers on the advancement of data-driven computational models aimed at improving predictive accuracy within this field.

To foster new predictive techniques, we aim to approximate a time-evolution representation of neurological diseases from ECoG data, through a version of the Krankheit-operator tailored (K-operator) [18], previously applied to fMRI-derived data of Parkinson's Disease [19] and Alzheimer-Perusini's Disease patients [20]. K acts on brain areas and functional connections to reproduce the effects of a neurological disease. The K-operator draws its conceptual foundation from historical accounts of external influences on neuropsychiatric phenomena [18], and the need for a unified view of several diseases through the analysis of brain-connectome alterations [21]. The pathways of anatomic, functional, and effective connectivity can indeed be altered by the presence of various diseases.

Here, we focus on a single epileptic subject, using K to capture information on temporal alterations in epilepsy.

This study seeks to expand the conceptual framework of the *K*-operator via a preliminary investigation of its dynamics, integrating empirical data from ECoG signals to model the temporal aspects of the disease.

The transition from the fMRI-based *K*-operator to the ECoG-based one is not trivial and specifically aims to investigate the correlation between the activity in different brain areas. Thus, our study constitutes a bridge between time-domain analysis and the correlation between the activity of different brain areas, in preparation for more comprehensive research relating activity within each area with the overall brain connectome.

2. Methods

2.1. Definition of the K-operator

The *K*-operator is conceptualized as an operator provoking alteration on a physical observable, thus it inherits the matrix and operatorial formalism from theoretical physics. Denoting by G the matrix of

weights of functional connections in a healthy brain, and G^k in a diseased brain, we define K as $KG=G^k$, i.e., K is a mathematical object that turns a healthy brain into a diseased one. Describing the time evolution of a diseased patient, K acts as $K(t)G^k(t)=G^k(t+1)$, describing the time evolution of the disease in a brain. Having the information on the brain matrices, we approximate K as $K(t)=G^k(t+1)G^k(t)^{-1}$. This equation is exactly solved via a proper matrix product. However, here we use a purely element-wise Hadamard product, limiting the idea of the matrix inversion to single elements (no entries are exactly zero). K is retrieved dividing each element of $G^k(t+1)$

$$K(t) * \mathcal{G}^{k}(t) = \mathcal{G}^{k}(t+1) \Rightarrow \{k(t)\}_{ij} = \frac{\{g_{ij}^{k}(t+1)\}}{\{g_{ij}^{k}(t)\}}, \quad (1)$$

where * denotes the element-wise product, $\{k(t)\}_{ij}$ is the *(ij)*-th element of the *K*-operator, and $\{g_{ij}^{k}(t+1)\},\$ $\{g_{ii}^{k}(t)\}\$ are the corresponding elements of matrices $G^{k}(t+1)$ and $G^{k}(t)$, respectively. As the brain matrices, we use the correlation matrices obtained from ECoG data. For the sake of simplicity, a linear definition is adopted. If we consider K as an element-wise multiplicative operator, then it can be considered as linear. However, the K defined with the inverse matrix is in general nonlinear; also, if we define K(t) as a differential operator, according to the specific disease, nonlinearity could be introduced. Thus, it is safer to not impose linearity upon the definition of K. In addition, as theoretically described in [18], K can act on different levels, including neuronal firing rate, interaction between neuronal populations, and interaction between lobes. Thus, K acts on the different layers of the brain network. In its current data applications, the high-level connectome approach has been privileged [19, 20], and the results obtained through K are confirmed by medical literature findings. Here, we start from regions of interest caught in fMRI to the inter-channel variation of neural activity, exploiting real data from an ECoG measurement to shape the form of the operator. Formally, in the case of fMRI, the matrix elements of G^k are computed as the correlation between the time series of pairs of regions of interest. Similarly, in the case of ECoG, the matrix elements of G^k are the correlation values between the signals (time series) of each pair of channels. The final correlation number between each pair of channels is computed by averaging the correlation values obtained in a specific time frame. We consider all channels for the computation of G^k . In Section 3, we show and discuss a submatrix of G^k corresponding to a selection of pairs of channels.

2.2. Algorithm

The proposed algorithm inputs consist of time-series data obtained from selected ECoG electrodes (or channels) and the specified number of

intervals (Fig. 1). First, we segment the ECoG channel data into *n* intervals (in our case, n=4). The correlation matrix between the channels is computed for each interval *w* (with w=1,...,n), denoted as G_w . Hence, the *K*-operator for each pair of consecutive time intervals *w* and w + 1 is represented as K(w,w+1). Then, K(t) is approximated via regression from the preceding elements.

2.3. Case Study

We consider as a case study the ECoG data collected on patient 02 of the Fragility Multi-Center Retrospective Study³. The patient is a woman of 28 y.o. with a hypothesized left anterior temporal lobe epilepsy. The ECoG electrodes of the resection surgery site are PST1-4, AST1-2, and MST1-2. Electrodes ALEX1-8, LAEX3, RQ1-2, G5-6, G17, and G25 are excluded from the analysis being bad channels, according to the annotations provided by the dataset expert. The first pre-surgery recording (Fig. 2) lasts 313.11s with a sampling rate of 1000 Hz. According to the annotations, an early onset of a seizure starts at the second 105.90, mainly seen on TT1 and then on the PST channels. The delta rhythm slows on G1-3 (109.78s), and a spread on the TT, AST, MST, and PST electrodes starts at 154.97s while a general spread starts at 164.85s. The epilepsy seizure offset is marked at 204.74s.

We focused on a single patient to test our methodology in the case of ECoG signals, before extending the analysis to multiple patients.

Algorithm 1: k-Operator Computation
Data: ECoG channels <i>ECoG_{Ch}</i> , # intervals <i>n</i>
Result: $K(t)$
for $c = 1, \dots, ECoG_{Ch}$ do
${ECoG_c^1,, ECoG_c^n} \leftarrow segment(ECoG_c, n);$
end
$\mathcal{G}_w \leftarrow \varnothing;$
for $w = 1,, n$ do
for $i = 1,, ECoG_{Ch} - 1$ do
for $j = i+1, \dots, ECoG_{Ch}$ do
$g_{wij} \leftarrow g_{wij} \leftarrow g_{w$
$Correlation(ECoG_i^w, ECoG_j^w);$
$G_{w}.append(g_{wij});$
end
end
end
for $w = 1,, n - 1$ do
$K_{w,w+1} \leftarrow rac{\mathcal{G}_{w+1}}{\mathcal{G}_w}$ /*see eq.1
end
$K(t) \leftarrow Regression(K_{(1,2)}, \dots, K_{(n-1,n)})$

Fig. 1. Algorithm.

3. Results and Discussions

Each ECoG signal (from the 76 electrodes) is divided into four intervals of 78.28 seconds each,

comprising different brain activations (Fig. 2) and compliant with the explanation given by the expert labeler of the data. The first window shows a normal-steady neural activity. The second one includes the early onset of the epileptic seizure. The third one encases the whole spread of the seizure and its offset. The fourth window presents the post-seizure neural activity. We compute the correlation matrices (our G^k s) between channels on each interval. Then, by looking at the K_{12} computed from $G^k(1)$, corresponding to the first interval, and $G^{k}(2)$, corresponding to the second interval (Fig. 3a), we notice that the pairs of channels TT2 - PST1 and TT2 - PST3 do not show a great correlation variation from the first to the second time interval. However, from K_{23} (Fig. 3b), we notice an increase in correlation variational trend for TT2 – TT1, TT2 – MST1, TT2 – PST1, TT2 – PST3. Finally, K_{34} (Fig. 3c) presents a more evident variation of TT2 - PST2, MST4 - PST3, PST4- MST2, PST4 - TT1, PST4 – PST1. This shows a correspondence between the time variations of the raw signals and the variations detected by the K-operator. Focusing on TT2, we notice an increase in signal variations with respect to the other channels. This may be due to the initial wide oscillation around the second 50. A similar seizure spreading is described in the literature [22], and temporal lobe epilepsy can have serious consequences, also involving language-network connectivity [23].

4. Conclusions

Our approach allows us to estimate the value of K(t) for every t. However, we stress that K(t) is computed via a regression from three instances of K from four intervals. Thus, part of the information is lost. As currently defined, the K-operator is more suitable for the interactions between brain areas rather than what happens inside a specific area.

Indeed, as currently defined for data applications, the K-operator is more suitable to describe the interactions between brain areas rather than what happens inside a specific area. In its present form, K(t)captures the "relative relationships" between channels, catching the electric activity between brain areas rather than the alteration of the activity in an area itself, which can characterize a seizure. Nevertheless, a seizure is also characterized by the synchronization between neuronal agglomerates in the same area or between different brain areas, thus, it is not only a "local" property. Considering the idea of writing K as the tensor product of its submatrices, we can think of a part of the operator that modifies the "inner behavior" of a certain area and multiplies as a tensor a group of brain areas. Alternatively, there could be a multi-layer K, with "inner layers." In this case, we could describe the layer of single channels as another layer.

In addition, a more refined approach would involve the computation of G^k for smaller windows and the

³ https://openneuro.org/datasets/ds003029/versions/1.0.6

regression to estimate K(t) directly from them, losing the minimal amount of information.

To take into account non-linearity, we also computed the approximation of K(t) via a quadratic

regression. We present two instances of K(t) as a methodological example at t=170s (Fig. 4a) and at t=230s (Fig. 4b).







Fig. 3. Ks between the four intervals, K_{12} , K_{23} , $K_{34...}$



Fig. 4. *K*(*t*) at specific times.

Here, our comments focused on trends, noticing the similarity with seizure spread as described in the literature. However, a more comprehensive discussion would require the comparison between different patients.

That said, observing the obtained values of elements of K, we caught some hints on the possible spreading of the seizure from the hippocampus through the temporal gyrus and ending within the superior temporal gyrus, as confirmed by the literature [22].

Possible further developments may involve the combination of the *K*-operator for the inter-channel correlation with a detailed study of intra-channel

information, via recurrence plots. This would help connect our research to the recurrence analysis of pre-ictal and inter-ictal periods from epileptic EEG data, and their measures of complexity [24].

Our final aim is tuning a prediction system, considering patients' variability, sex, and age, predicting onset and evolution of seizures and key features of seizure propagation.

Code. https://github.com/medusamedusa/K-operator_epilepsy

Funding. "Age-It – Ageing well in an ageing society" project (PE0000015), National Recovery and Resilience Plan (NRRP) – PE8 – Mission 4, C2, Intervention 1.

Acknowledgments

We thank Gemma Alfano, linguist and professor of Latin and Greek, for her explanation of etymology and its connection with key features of the disease.

References

- S. J. Baloyannis, Epilepsy: A Way from Herodotus to Hippocrates, *Epilepsy & Behavior*, Vol. 28, Issue 2, 2013, 303.
- [2]. C. Lainscsek, P. Salami, V. R. Carvalho, E. M. A. M. Mendes, M. Fan, S. S. Cash, T. J. Sejnowski, Networkmotif delay differential analysis of brain activity during seizures, *Chaos*, Vol. 33, Issue 12, 2023, 123136.
- [3]. J. Royer, B. C. Bernhardt, S. Larivière, E. Gleichgerrcht, B. J. Vorderwülbecke, S. Vulliemoz, L. Bonilha, Epilepsy and brain networks hubs, *Epilepsia*, Vol. 63, 2022, 537.
- [4]. F. Bartolomei, M. Guye, F. Wendling, Abnormal binding and disruption in large scale networks involved in human partial seizures, *EPJ Nonlinear Biomedical Physics*, Vol. 1, Issue 4, 2013, 1.
- [5]. X. Zhu, H. Shappell, M. Kramer, C. Chu, E. Kolaczyk, Distinguishing between different percolation regimes in noisy dynamic networks with an application to epileptic seizure, *PLoS Computational Biology*, Vol. 19, Issue 6, 2023, e1011188.
- [6]. R. Akut, Wavelet-based deep learning approach for epilepsy detection, *Health Information Science and Systems*, Vol. 8, Issue 7, 2019, 1.
- [7]. P. Barone, et al., Neurologia Clinica, *Idelson-Gnocchi*, Naples, 2021.
- [8]. A. F. Jackson, D. J. Bolger, The neurophysiological bases of EEG and EEG measurement: a review for the rest of us, *Psychophysiology*, Vol. 51, 2014, pp. 1061-1071.
- [9]. Z. Wang, X. Song, L. Chen, J. Nan, Y. Sun, M. Pang, K. Zhang, X. Liu, D. Ming, Research progress of epileptic seizure prediction methods based on EEG, *Cognitive Neurodynamics*, Vol. 18, 2024, pp. 2731-2750.
- [10]. B. Direito, C. A. Teixeira, F. Sales, M. Castelo-Branco, A. Dourado, A realistic seizure prediction study based on multiclass SVM, *International Journal of Neural Systems*, Vol. 27, Issue 3, 2017, 1750006.

- [11]. Y. Sun, W. Jin, X. Si, X. Zhang, J. Cao, L. Wang, S. Yin, D. Ming, Continuous seizure detection based on transformer and long-term iEEG, *IEEE Journal of Biomedical and Health Informatics*, Vol. 26, Issue 11, 2022, pp. 5418–5427.
- [12]. M. Savadkoohi, T. Oladunni, L. Thompson, a machine learning approach to epileptic seizure prediction using Electroencephalogram (EEG) signal, *Biocybernetics* and Biomedical Engineering, Vol. 40, Issue 3, 2020, pp. 1328-1341.
- [13]. Z. Yu, W. Nie, W. Zhou, F. Xu, S. Yuan, Y. Leng, Q. Yuan, Epileptic seizure prediction based on local mean decomposition and deep convolutional neural network, *The Journal of Supercomputing*, Vol. 76, 2020, pp. 3462-3476.
- [14]. J.-E. Le Douget, A. Fouad, M. M. Filali, J. Pyrzowski, M. Le Van Quyen, Surface and intracranial EEG spike detection based on discrete wavelet decomposition and random forest classification, in *Proceedings of the 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (*EMBC'17*), Jeju Island, South Korea, 11-15 July 2017, pp. 475-478.
- [15]. G. Varotto, L. Tassi, S. Franceschetti, R. Spreafico, F. Panzica, Epileptogenic networks of type II focal cortical dysplasia: a stereo-EEG study, *NeuroImage*, Vol. 61, Issue 3, 2012, pp. 591-598.
- [16]. S. B. Tomlinson, B. E. Porter, E. D. Marsh, Interictal network synchrony and local heterogeneity predict epilepsy surgery outcome among pediatric patients, *Epilepsia*, Vol. 58, Issue 3, 2017, pp. 402-411.
- [17]. S. S. Balaji, K. K. Parhi, Seizure onset zone identification from iEEG: a review, *IEEE Access*, Vol. 10, 2022, pp. 62535-62547.
- [18]. M. Mannone, P. Fazio, N. Marwan, Modeling a neurological disorder as the result of an operator acting on the brain: a first sketch based on network channel modeling, *Chaos*, Vol. 34, Issue 5, 2024, 053133.
- [19]. M. Mannone, P. Fazio, J. Kurths, P. Ribino, N. Marwan, A brain network operator for modeling disease: a first data-based application for Parkinson's disease, *European Physical Journal, Special Topics*, 2024.
- [20]. M. Mannone, N. Marwan, P. Fazio, P. Ribino, Limbic and cerebellar effects in Alzheimer-Perusini's disease: a physics-inspired approach, *Biomedical Signal Processing and Control*, Vol. 103, 2025, 107355.
- [21]. W. W. Seeley, Mapping neurodegenerative disease onset and progression, *Cold Spring Harbor Perspectives in Biology*, Vol. 9, 2017, a023622.
- [22]. E. Gleichgerrcht, A. S. Greenblatt, T. S. Kellermann, N. Rowland, A. V. W., J. Edwards, K. A. Davis, L. Bonilha, Patterns of seizure spread in temporal lobe epilepsy are associated with distinct white matter tracts, *Epilepsy Research*, Vol. 171, 2021, 106571.
- [23]. K. Trimmel, A. L. van Graan, L. Caciagli, A. Haag, M. J. Koepp, P. J. Thompson, J. S. Duncan, Left temporal lobe language network connectivity in temporal lobe epilepsy, *Brain*, Vol. 141, 2018, pp. 2406–2418.
- [24]. E. J. Ngamga, S. Bialonski, N. Marwan, J. Kurths, C. Geier, K. Lehnertz, Evaluation of selected recurrence measures in discriminating pre-ictal and inter-ictal periods from epileptic EEG data, *Physics Letters A*, Vol. 380, 2016, pp. 1419-1425.

(032)

Millimeter-wave Beam Prediction with Inverse Beamforming ML Model

S. Mokdadi, S. E. Bouzid and P. Chargé

Nantes University, CNRS, IETR, UMR 6164, Rue Christian Pauc, F-44306 Nantes, France Tel.: +33 240683242 E-mail: smail.mokdadi@etu.univ-nantes.fr

Summary: Next-generation wireless systems rely on millimeter-wave (mmWave) frequencies for high bandwidth, but their propagation is challenged by path loss and environmental blockages, affecting reliability. Beam management, particularly beam steering, is essential to maintaining robust communication in dynamic environments. Traditional methods estimate the current Angle of Arrival (AoA), which struggle with high user mobility scenarios, leading to increased connection outages. In contrast, predicting the future AoA enables proactive beam adjustments, reducing delays and improving link stability. This work focuses on AoA prediction at the UE side with limited resources, ensuring real-time operation. We propose a low-complexity LSTM-based model that predicts future AoA using prior ones and channel observations, requiring only the beamformer's output signal. Experimental results demonstrate that our ML-based solution significantly reduces outage probability compared to the EKF, particularly in low-SNR conditions, highlighting its effectiveness in dynamic mmWave environments.

Keywords: Angle of arrival (AoA), Extended Kalman filter (EKF), Long short-term memory (LSTM), Machine learning, Millimeter-wave (mmWave).

1. Introduction

Next-generation wireless systems rely on millimeter-wave (mmWave) frequencies for high bandwidth, however, propagation challenges such as severe path loss and environmental blockages can compromise reliability. Effective beam management including beam search for link establishment or recovery and beam steering for dynamic adjustments is therefore essential in these dynamic environments.

Accurate and timely prediction of the Angle of Arrival (AoA) is crucial for efficient beam steering. Traditional methods depend on beam sweeping to estimate the AoA, a process that demands extensive pilot signals and significant radio resources. This increases latency due to the time required for measurement and data processing, thereby elevating the risk of connection outages. In contrast, tracking techniques that operate solely on the received signal avoid additional radio resource consumption.

Our solution addresses these challenges by employing a machine learning (ML) approach that leverages a Long Short-Term Memory (LSTM) network for beam tracking at the user equipment (UE) side, where resources, energy, and real-time processing capabilities are limited. This method predicts the AoA for the next time step by analyzing prior AoA values and channel observations, enabling the UE to anticipate changes in beam direction and ensuring seamless alignment and reduced latency.

The remainder of this paper is organized as follows: Section 2 reviews related works. Section 3 introduces the system model, formulates the problem, and describes the data generation process; Section 4 presents the EKF- and ML-based solutions; Section 5 provides a performance analysis and comparison of the proposed solutions; and Section 6 concludes the paper.

2. Related Works

Recent studies have explored various approaches for beam tracking. Several works have employed Extended Kalman Filter (EKF) techniques to address challenges such as beam tracking under line-of-sight (LOS) conditions [1] and handling multipath channels dominated by a single LOS path [2]. EKF methods have also been applied in V2X systems for intersection management [3] and precise tracking of vehicle position and motion [4]. Moreover, adaptive EKF approaches have been developed to balance accuracy and overhead in multipath environments [5]. However, the EKF solution's limited generalization capacity under dynamic conditions and its reliance on numerous parameters impose additional constraints that complicate practical implementation. Consequently, in our study, the EKF is used primarily as a benchmark.

In parallel, various ML approaches have been investigated for beam prediction in mmWave systems [6]. Some studies have compared algorithms such as KNN, SVM, decision trees, and naïve Bayes for beam prediction accuracy [7]. Other research has explored encoder-decoder architectures using visual sensing for future beam prediction in V2X data communications [8], while low-complexity ML designs exploiting RSRP have also been introduced [9]. Notably, LSTM-based models have gained significant traction, with applications in predicting channel behavior [10], tracking AoA using prior channel observations [11], and determining optimal serving beams from beam index sequences [12]. Although these ML approaches effectively capture temporal dependencies, most focus on estimating the current AoA or require high-complexity models.

Building on these insights, our proposed ML-based solution leverages an LSTM network within a

low-complexity model to predict the AoA for the next time step. This proactive strategy enables the user equipment (UE) to anticipate changes in beam direction, ensuring seamless beam alignment and improved reliability in high-mobility scenarios.

The contributions of this paper are twofold:

• First, it presents a low-complexity LSTM-based model that predicts AoA in mmWave systems using only the observed signal from the beamformer output. This approach optimizes the balance between computational efficiency and prediction accuracy for real-time beam steering.

• Second, the proposed model demonstrates robust adaptability by generalizing across various antenna configurations and user mobility conditions. Even when trained on a single speed or specific antenna setup, it maintains high performance without the need for retraining, offering a scalable solution that performs well in diverse and dynamic scenarios.

3. System Model and Data Generation

In this section, we present the system model for beam direction prediction. We consider a scenario where an omni-directional BS communicates with UEs equipped with a uniform linear array (ULA) of N antenna elements. Equation (1) defines the steering vector characterizing the ULA's response [1, 2, 9, 10].

$$\boldsymbol{a}(\theta) = \left[1, e^{jkd\cos(\theta)}, \dots, e^{jkd(N-1)\cos(\theta)}\right]^{T}, \quad (1)$$

where $k = 2\pi/\lambda$, λ is the propagation signal wavelength, *d* is the distance between adjacent antenna elements, θ is the true AoA, and *N* is the antenna number at the UE. The channel observation at a given time *t* is described by:

$$y[t] = \frac{\alpha[t]}{N} \mathbf{w}^{H}(\hat{\theta}[t]) \mathbf{a}(\theta[t]) + n[t]$$
(2)

A single dominant multipath component is assumed to be tracked and predicted, as secondary paths lie outside the main beam steered by the antenna array. Here, $\alpha[t]$ denotes the complex gain of the dominant path, $\boldsymbol{a}(\theta[t])$ represents the steering vector, and $\boldsymbol{w}(\hat{\theta}[t])$ is the beamforming weighting vector based on the predicted angle. The notation $(.)^H$ denotes the Hermitian operator (complex conjugate transpose), n[t] represents additive noise and interference, finally $\theta[t]$ and $\hat{\theta}[t]$ denote the true and predicted AoA respectively.

In addition to this observation model, we assume that an initial connection between the BS and UE has been established, with the main challenge being to maintain the connection via beam tracking. At t = 0, the model requires prior AoA values to ensure sufficient historical data for predictions. Although simplified for clarity, our models can be extended to scenarios where the BS uses an antenna array.

The evolution processes for both the real and imaginary part of the path gain are assumed to follow the first-order Gauss-Markov model given by [1, 2]:

$$\alpha[t+1] = \rho \alpha[t] + \zeta[t], \qquad (3)$$

where ρ is the correlation coefficient and $\zeta[t]$ is a random variable with a normal distribution $\zeta[t] \sim \mathcal{N}(0, (1 - \rho^2)/2)$.

To generate data for this study, we simulated a mmWave communication scenario in which a BS is positioned at the center of a cell. User movement is modeled with varying speed and orientation based on predefined laws to ensure realistic mobility. These characteristics, especially speed and direction, significantly affect beam prediction performance. Our models incorporate gradual changes in these variables, with current values influenced by past states to ensure a smooth and realistic evolution. The simulation parameters are summarized in Table 1.

Table 1. Simulation Parameters.

Parameter	Value
Cell Radius	60 meters
Carrier Frequency	28 GHz
BS Antenna	Omnidirectional
UE Antenna	ULA / N \in {4, 8, 12,, 64}
Prediction period	0.1 seconds
α Model	First-Order Gauss-Markov
Users speed	speed $\in [1, 33]$ m/s
Users orientation	Varying orientation

4. Proposed Solutions

In this section, we present two approaches for predicting the AoA at the next time step. First, we introduce an EKF-based solution as a benchmark for its effectiveness in addressing nonlinear estimation problems. Next, we detail our ML-based approach, which leverages an LSTM network.

4.1. EKF Solution

The EKF-based solution offers low computational complexity and reliable performance suitable for real-time beam prediction. However, its reliance on numerous parameters complicates the practical implementation and limits adaptability under dynamic conditions. The state evolution model is defined using a state vector comprising the real and imaginary parts of the path gain and the AoA based on the variables proposed in [2]. This state vector is given in (4), and the AoA evolution, represented by a Gaussian process noise model, is expressed in (5) [5].

$$\boldsymbol{x}[t] = \left[\Re(\boldsymbol{\alpha}[t]), \Im(\boldsymbol{\alpha}[t]), \boldsymbol{\theta}[t]\right]^{T}, \quad (4)$$

$$\theta[t+1] = \theta[t] + \zeta_{\theta}[t], \qquad (5)$$

where $\zeta_{\theta}[t]$ is a random variable distributed according to $\zeta_{\theta}[t] \sim \mathcal{N}(0, \sigma_{\theta}^2)$ and σ_{θ}^2 is the angle variance in one time slot. The state evolution model, according to the discrete-time stochastic evolution model, is expressed as follows [1, 2]:

$$\boldsymbol{x}[t+1] = \boldsymbol{F}\boldsymbol{x}[t] + \boldsymbol{u}[t] + \boldsymbol{w}[t]$$
(6)

From (3) and (5), the state transition matrix F is given by $F = diag([\rho, \rho, 1])$ where diag(a) denotes the diagonal matrix with diagonal elements from the vector a. The control input u[t] represents an external influence on the state evolution and it's defined as u[t] = x[t] - x[t - 1], this formulation models the state variation, assuming that for small time steps, the variation remains approximately constant. The process noise w[t] models uncertainties in the state evolution, and defined as $w[t] \sim \mathcal{N}(0, W)$ where W is the process noise covariance matrix, given by $W = [(1 - \rho^2)/2, (1 - \rho^2)/2, \sigma_{\theta}^2]$.

The objective of the EKF is to recursively update the previous state prediction $\mathbf{x}[t]$ using the observed signal y[t], as defined in (2), and the measurement model $h(\mathbf{x}[t])$. Here, $h(\mathbf{x}[t])$ represents the observation function, which depends on three variables extracted from $\mathbf{x}[t]$ and it is defined as y[t] in (2), but without the noise component n[t] [1, 2]. Subsequently, the EKF predicts the channel parameters $\mathbf{x}[t+1]$ for the next time step. The EKF algorithm for one time step is presented as follows [1]:

Initialization:
$$\tilde{\mathbf{x}}[t-1], \mathbf{y}[t], \tilde{\mathbf{\Sigma}}[t-1]$$

 $\widehat{\mathbf{x}}[t] = F\tilde{\mathbf{x}}[t-1] + \mathbf{u}[t-1]$
 $\widehat{\mathbf{\Sigma}}[t] = F\tilde{\mathbf{\Sigma}}[t-1]F^T + W$
Loop $(t \leftarrow t+1)$:
• Kalman Gain
 $K[t] = \widehat{\mathbf{\Sigma}}[t]H^T[t](H[t]\widehat{\mathbf{\Sigma}}[t]H^T[t] + V)^{-1}$
• State Update
 $\tilde{\mathbf{x}}[t] = \widehat{\mathbf{x}}[t] + K[t](\mathbf{y}[t] - h(\widehat{\mathbf{x}}[t]))$
 $\widetilde{\mathbf{\Sigma}}[t] = (I - K[t]H[t])\widehat{\mathbf{\Sigma}}[t]$
• State Prediction
 $\widehat{\mathbf{x}}[t+1] = F\tilde{\mathbf{x}}[t] + \mathbf{u}[t]$
 $\widehat{\mathbf{\Sigma}}[t+1] = F\tilde{\mathbf{\Sigma}}[t]F^T + W$

The predicted channel parameters $\hat{x}[t+1]$ and the predicted error covariance $\hat{\Sigma}[t+1]$ are determined based on the state evolution model, the previous updated estimates $\check{\mathbf{x}}[t]$ and $\check{\mathbf{\Sigma}}[t]$. Here, K[t] denotes the Kalman gain, I is the identity matrix, and H[t] is the Jacobian of $h(\mathbf{x}[t])$ with respect to the state vector x[t] [2]. To ensure that the state is represented in real values, the observed signal is expressed as $\mathbf{y}[t] = [\Re(\mathbf{y}[t]), \Im(\mathbf{y}[t])]^T$ [2] and the measurement error covariance matrix is defined as $\mathbf{V} = diag([1/(N.SNR), 1/(N.SNR)]).$ It is important to note that the EKF algorithm requires knowledge of the SNR during testing to accurately generate this matrix, which introduces an additional constraint.

4.2. ML Solution

Our ML-based solution employs an LSTM network chosen for its ability to capture long-term dependencies in time-series data, allowing the model to leverage multiple prior states from previous AoA estimates and channel observations for accurate predictions of AoA at the next time step. Fig. 1 illustrates our proposed approach, where the model is implemented as an LSTM network block comprising a fully connected layer with 30 units, followed by an LSTM layer with 50 hidden units, and a final fully connected layer with 30 units. No activation functions are applied to the output, and the layer sizes were determined empirically to balance model capacity and computational efficiency.

The network input comprises noisy observed signals $\{y[t-1], y[t-2], ..., y[t-L]\}$ from the beamformer output, as defined in (2), along with $\{\hat{\theta}[t-1], \hat{\theta}[t-1]\}$ AoA predictions previous 2], ..., $\hat{\theta}[t-L]$. This setup reflects the practical scenario in which the UE has access only to predicted AoA values and measured v. The sequence length L determines the number of prior states considered. Our model effectively performs an inverse beamforming process on the observed signals to extract the AoA. By analyzing the sequential pattern of past AoA values, it learns the underlying variation patterns and temporal dependencies, enabling it to predict the next AoA $\hat{\theta}[t+1]$. To ensure efficient training, the model is optimized using the Adam optimizer.



Fig. 1. The proposed beam prediction approach.

In our model, the parameter *L* defines the sequence length, representing the number of prior states included as input. For instance, when L = 2, the input includes both the previous state and the one before it. Selecting an optimal sequence length is crucial to balance sufficient information for accurate predictions without overwhelming the network with excessive complexity, as we demonstrate in the results section.

In practical scenarios, previous predictions affect channel observations because they determine the beam directions. To simulate this during training and ensure that our model is robust against uncertainty inherent in
prior predictions, we define the training AoA as $\theta_{trn}[t] = \theta[t] + \psi[t]$ [11], where $\psi[t]$ is a zero-mean Gaussian noise with a standard deviation of 4°, reflecting typical prediction errors. The corresponding channel observation $y_{trn}[t]$ is then generated based on both the true AoA $\theta[t]$ and the training AoA $\theta_{trn}[t]$. During training, the target label is the true AoA at the next time step $\theta[t + 1]$, while the input features include the real and imaginary components of the previous observations y_{trn} along with the training previous AoA values θ_{trn} .

During testing, an initial sequence of historical AoA values and channel observations (of length L) is assumed to be known, ensuring the model has sufficient data for predictions. Thereafter, the model sequentially predicts each new AoA using its prior outputs and the corresponding channel observations. Note that the channel observation y[t] depends on both the UE's beam orientation decision $\hat{\theta}[t]$ and the true AoA $\theta[t]$.

5. Results

In this section, we evaluate the performance of the above-presented beam prediction approaches, the EKF and our proposed ML solution, using two key metrics: outage probability and Root Mean Squared Error (RMSE). At each time step, the model predicts the AoA $\hat{\theta}$ and calculates the error with respect to the true AoA θ . If the prediction error exceeds a predefined threshold $\Delta \theta_{Th}$, the connection is considered lost, and an outage is recorded. For simplicity, we assume that the connection is reestablished instantaneously, and the error for that step is excluded from the RMSE calculation since it represents a connection failure rather than a prediction error.

The outage probability quantifies the likelihood of a connection loss when the predicted AoA deviates from the true AoA by more than the threshold, and is defined as follows:

$$P = \frac{1}{M} \sum_{m=1}^{M} I[\left|\theta_m - \hat{\theta}_m\right| > \Delta \theta_{Th}], \qquad (8)$$

$$\Delta\theta_{Th} = \frac{4\pi}{3N},\tag{9}$$

where, θ_m and $\hat{\theta}_m$ represent the true and predicted AoA at the *m*-th instance, respectively, *M* represents the total number of predictions, *I*[.] is an indicator function that equals 1 if its argument is true and 0 otherwise, and the threshold angle $\Delta \theta_{Th}$ is determined as a function of the UE antennas [11]. A lower outage probability reflects improved link reliability, directly contributing to a better user experience.

The RMSE (Root Mean Squared Error) quantifies the accuracy of the predicted AoA $\hat{\theta}$ by measuring the average magnitude of the error between the true AoA and its prediction. It is calculated as:

$$RMSE = \sqrt{\frac{1}{K} \sum_{k=1}^{K} \left(\theta_k - \hat{\theta}_k\right)^2},$$
 (10)

where θ_k and $\hat{\theta}_k$ are the true and predicted AoA at the *k*-th instance (for non-outage cases), and *K* is the total number of such non-outage instances. A lower RMSE indicates higher prediction accuracy.

Fig. 2 illustrates the outage probability and RMSE as functions of the sequence length L for different Signal-to-Noise Ratio (SNR) values at a mean user speed of 1 m/s. The results show that increasing the sequence length generally improves performance by providing the model with a richer historical context to capture user mobility and channel dynamics. However, when L exceeds 4, the performance degrades, as the excess data overwhelms the model's capacity to process it effectively, leading to diminishing gains.



Fig. 2. Outage probability and RMSE as a function of Sequence Length for ML solution.

Based on these observations, we select a sequence length L = 3 for the remainder of the results, balancing sufficient historical information with manageable data complexity and prediction accuracy.

Fig. 3 illustrates the outage probability and RMSE as functions of SNR for the ML solution, with curves corresponding to different training SNR values. In the upper part, we observe that low-SNR trained models perform better in low SNR scenarios. This is evident from the lower outage probabilities achieved at poor SNR levels. This outcome can be attributed to the model's ability to generalize and adapt when trained in more challenging conditions, making it more robust in environments with high noise levels. In the lower part, we notice that models trained at lower SNRs exhibit lower accuracy in high-SNR scenarios compared to those trained at higher SNRs. This is likely because high-SNR-trained models are better optimized for low-noise environments, while low-SNR-trained models prioritize robustness over precision.

Based on these observations, we select SNR training=-2 dB for the remainder of the results. This choice represents a trade-off, balancing the model's performance across both low and high SNR conditions.

7th International Conference on Advances in Signal Processing and Artificial Intelligence (ASPAI' 2025), 8-10 April 2025, Innsbruck, Austria



Fig. 3. Outage probability and RMSE as a function of SNR for ML solution with different training SNR.

Fig. 4 illustrates the performance comparison between the proposed ML solution with a sequence length L = 3 and the EKF across varying SNR levels with a mean user speed of 1 m/s.



Fig. 4. ML vs. EKF: Outage probability and RMSE as a function of SNR.

In terms of outage probability, the ML solution consistently outperforms the EKF across all SNR values, demonstrating its superior ability to maintain reliable connections under challenging conditions. However, at lower SNR levels, the ML model exhibits a higher RMSE than the EKF because it prioritizes connection reliability over prediction precision. In contrast, the EKF maintains a lower RMSE as it does not account for outage instances in its error calculations. As SNR increases, both methods show a reduction in RMSE, eventually converging to similar values, which indicates comparable prediction accuracy in favorable conditions. Overall, these observations highlight the ML solution's superior capability in reducing outage probability.

Fig. 5 illustrates the outage probability and RMSE as functions of the users' mean speed at an SNR of 0 dB. The analysis compares three approaches: the ML solution with L = 3 trained at a speed of 10 m/s (Single training), the ML solution retrained for each specific speed with L = 3 (Separate training), and the EKF method.



Fig. 5. Outage probability and RMSE as a function of speed.

Both ML approaches outperform the EKF method, demonstrating their effectiveness in AoA prediction. While the ML solution with separate training for each speed performs slightly better than the single training model, the small difference highlights the ML model's ability to generalize across different mobility conditions. However, further performance gains can be achieved by tailoring the training process to specific scenarios, enhancing robustness against high mobility challenges. Regarding RMSE, all methods exhibit a similar trend: prediction errors increase with higher user speeds. This indicates that as users move faster, accurately predicting the Angle of Arrival (AoA) becomes more challenging for all approaches.

Fig. 6 illustrates the outage probability and RMSE as functions of users' antenna number at an SNR of 0 dB. The analysis compares three approaches: the ML solution with L = 3 trained with 16 antennas (Single Train), the ML solution retrained for each antenna configuration (Separate Train), and the EKF method.



Fig. 6. Outage probability and RMSE as a function of the number of antennas.

Both ML approaches outperform the EKF method, demonstrating superior AoA prediction capabilities. Although the ML solution with separate training for each antenna configuration performs slightly better than the single training model, the small difference highlights the model's strong generalization across different setups. However, further performance improvements can be achieved by tailoring the training process to specific antenna configurations. Regarding RMSE, all methods show a similar trend: prediction errors decrease as the number of antennas increases due to a smaller $\Delta \theta_{Th}$. In addition to performance metrics, we compare the computational complexity of the ML and EKF methods in terms of floating-point operations (FLOPs). The ML solution with L = 3requires approximately 32,650 FLOPs per prediction step, significantly lower than many existing ML-based approaches, while the EKF requires only 774 FLOPs. However, due to its higher outage probability, the EKF method necessitates at least eight times more beam searches to compensate for outages, leading to substantial radio resource consumption. This highlights the efficiency of our ML solution, which, despite its higher computational cost, minimizes beam searches and offers a well-balanced trade-off between efficiency, reliability, and adaptability.

6. Conclusions

In this paper, we introduced a novel LSTM-based approach for predicting beam direction in mmWave systems. Our experiments demonstrate that the proposed ML method significantly reduces outage probability, especially under low SNR conditions, enhancing the reliability of mmWave links compared to the traditional EKF. We also emphasized the importance of selecting an optimal sequence length to balance sufficient information with accurate predictions. Furthermore, our approach shows robust adaptability, performing strongly even when trained on a single configuration, however tailoring the training process to specific configurations or conditions can further enhance robustness. Overall, our study highlights the potential of ML techniques, particularly LSTM networks, in advancing beam management and AoA predictions. By enhancing beam steering reliability, our method reduces the need for frequent beam searches, paving the way for more efficient and adaptive mmWave communication networks.

References

 L. Chen, S. Zhou, W. Wang, MmWave beam tracking with spatial information based on extended Kalman filter, *IEEE Wirel. Commun. Lett.*, Vol. 12, Issue 4, 2023, pp. 615-619.

- [2]. M. Al-Ibadi, F. E. Mahmood, Beam and channel tracking for 5G communication systems using adaptive filtering techniques: a comparison study, *J. Commun. Softw. Syst.*, Vol. 18, Issue 3, 2022, pp. 244-251.
- [3]. D. Hu, J. Nakazato, E. Javanmardi, M. Asad, M. Tsukada, An extended Kalman filter enabled beam tracking framework in intersection management, in *Proceedings of the Eur. Conference Netw. Commun.* (EuCNC 6G Summit'23), Poznan, Poland, 3-6 June 2023.
- [4]. X. Meng, F. Liu, C. Masouros, W. Yuan, Q. Zhang, Z. Feng, Vehicular connectivity on complex trajectories: roadway-geometry aware ISAC beam-tracking, *IEEE Trans. Wirel. Commun.*, Vol. 22, Issue 11, 2023, pp. 7408-7423.
- [5]. M. Mo, C. Liu, L. Zhao, M. Guo, M. Li, Y. Wang, Iterated extended Kalman filter based adaptive beam tracking for millimeter-wave systems, in *Proceedings* of the IEEE/CIC International Conference on Communications in China (ICCC'24), Hangzhou, China, 7-9 August 2024, pp. 1585-1590.
- [6]. W. Yi, W. Zhiqing, F. Zhiyong, Beam training and tracking in mmWave communication: A survey, *China Commun.*, Vol. 21, Issue 6, 2024, pp. 1-22.
- [7]. K. K. Biliaminu, S. A. Busari, J. Rodriguez, F. Gil-Castiñeira, Beam prediction for mmWave V2I communication using ML-based multiclass classification algorithms, *Electronics*, Vol. 13, Issue 13, 2024, 2656.
- [8]. S. Jiang, A. Alkhateeb, Computer vision aided beam tracking in a real-world millimeter wave deployment, in *Proceedings of the IEEE Globecom Workshops (GC Wkshps'22)*, Rio de Janeiro, Brazil, 4-8 December 2022, pp. 142-147.
- [9]. M. Q. Khan, A. Gaber, M. Parvini, P. Schulz, G. Fettweis, A low-complexity machine learning design for mmWave beam prediction, *IEEE Wirel. Commun. Lett.*, Vol. 13, Issue 6, 2024, pp. 1551-1555.
- [10]. S. H. Lim, S. Kim, B. Shim, J. W. Choi, Deep learningbased beam tracking for millimeter-wave communications under mobility, *IEEE Trans. Commun.*, Vol. 69, Issue 11, 2021, pp. 7458-7469.
- [11]. D. Burghal, N. A. Abbasi, A. F. Molisch, A machine learning solution for beam tracking in mmWave systems, in *Proceedings of the 53rd Asilomar Conference on Signals, Systems, and Computers,* Pacific Grove, CA, USA, 3-6 November 2019, pp. 173-177.
- [12]. A. O. Kaya, H. Viswanathan, Deep learning-based predictive beam management for 5G mmWave systems, in *Proceedings of the IEEE Wireless Communications and Networking Conference* (WCNC'21), Nanjing, China, 29 March - 1 April 2021, pp. 1-7.

(033)

Video-based Analysis for Automated Ptosis Detection

S. Baliński and <u>P. Śniatała</u>

Poznan University of Technology, pl. M. Sklodowskiej-Curie 5, 60-965 Poznan, Poland E-mail: pawel.sniatala@put.poznan.pl

Summary: Myasthenia gravis (MG) belongs to the group of rare diseases. The development of a computer-based tool to facilitate the diagnosis of MG and collect data from clinical trials for further analysis is an important element that supports the development of MG treatment. This paper presents an original solution for supporting the diagnosis of a disease called Myasthenia Gravis (MG). Due to the slow progression of this disease, an important requirement of this system was the ability for the patient to self-monitor their condition at home. In particular, here we focus on an algorithm which automatically detects ptosis symptoms based on patient video captures with the standard camera. The predictor's ability to identify ptosis, clinically referred to as drooping of the eyelid, requires a sequential image analysis methodology. Each phase within this methodology is tasked with different visual data processing objectives that cumulatively culminate in the determination of the presence of ptosis. This system offers a low-cost and accessible alternative.

Keywords: Ptosis, Myasthenia gravis, Face detection, Eyelid detection.

1. Introduction

Myasthenia gravis (MG) belongs to the group of rare diseases. Rare diseases are most often genetically determined, with a chronic and often severe course, about half of which manifest themselves in childhood. Due to their rarity, difficulty in diagnosing, and lack of public awareness, knowledge about these diseases has been limited to date. The development of a computerbased tool to facilitate the diagnosis of MG and collect data from clinical trials for further analysis is an important element that supports the development of MG treatment [1].

In this article, we present a solution that allows automatic detection of one of the symptoms of MG, namely ptosis. The presented algorithm is a component of the DIAG-MG system developed by our team, which assesses four symptoms of MG: eyelid drooping, double vision, dysphagia and upper limb muscle strength [2]. Due to the slow progression of this disease, an important requirement of this system was the ability for the patient to self-monitor their condition at home. Before symptom assessment, patients are self-assessed using the Myasthenia Gravis Activities of Daily Living (MG-ADL) scale, thus establishing an initial baseline of symptomatology.

The topic related to an automated eyelid measurement was presented in some publications. Recently approaches using neural network technology have been utilized for this task. Publication [3] evaluates the clinical usefulness and reliability of a NN-based automated eyelid measurement system. Authors proposed an automated NN-based measurement system that could provide straightforward and precise method for measuring MRD1 and MRD2, as well as detecting morphological abnormalities in the eyelids. Another approach using NN was presented in [4]. Authors trained a neural network for eye landmark detection consisting of a ResNet50 backbone. They proved the feasibility of automated ptosis assessment from frames of video data collected remotely over a broad range of smartphones.

The solution presented in our proposal is based on the standard (non-NN) approach. Our system does not require a lot of computing power and can be run on devices that do not have a lot of resources. It is also possible to run this application on a smartphone.

2. DIAG-MED System

Elaborated by our team DIAG-MG system is a software, facilitates the assessment of symptom manifestation in individuals diagnosed with Myasthenia Gravis (MG) and quantifies the severity of these symptoms (severe, moderate, mild, or absence of symptom) in alignment with the criteria set forth by the Quantitative Myasthenia Gravis Test.

The DIAG-MG program is designed to evaluate four specific symptoms: ptosis, diplopia, dysphagia, and the muscular strength of the upper limbs. Prior to the symptomatic evaluation, patients will undergo a self-assessment utilizing the Myasthenia Gravis Activities of Daily Living (MG-ADL) scale, thereby establishing a preliminary baseline of symptomatology.

3. Ptosis Detection Algorithm

The prediction analysis component of DIAG-MG (predictor) is a sophisticated application element that uses various algorithms and image processing techniques to analyze visual data collected during diagnostic assessments. Its main task is to autonomously assess patient video recordings in order to identify and quantify specific neurologically relevant symptoms. After completion of the recording phase of the assessment, all video data is transmitted to the prediction software for processing.

The performance of the Predictor module in the identification of ptosis [5-7], or drooping of the eyelids, is achieved by a multistep image analysis process. Each phase is responsible for specific visual data processing tasks, which, in total, result in determining whether a patient has ptosis. The process includes the following six steps: Face Detection, Image Clipping for the Left and Right Eyes, Filter Application and Information Extraction, Pupil Search, and Eyelid Search.

3.1. Face Detection

In the first stage of the ptosis identification process, face detection plays a key role, for which Predictor uses the MediaPipe module. MediaPipe is an advanced solution developed by Google that enables the detection of faces and their key landmarks (landmarks) in images and videos. MediaPipe uses deep learning algorithms to efficiently and accurately detect faces in images. This allows it to quickly locate faces within the recorded video. After face detection, MediaPipe identifies key landmarks such as eyes, nose, mouth, and facial contour. These landmarks are essential for precise image cropping and further analysis of specific parts of the face, particularly the eye area.

3.2. Image Clipping for Left and Right Eyes

The next step in the process of identifying ptosis is to precisely cut the image into separate areas for each eye. To do this, Predictor uses facial landmarks obtained from the MediaPipe module, focusing on specific landmarks for the right and left eyes.

The Fig. 1 presents the example facial landmarks and the Fig. 2 shows an landmarks selection. Landmarks with IDs 53 and 233 are used for the right eye (shows in the Fig. 2), while 283 and 453 are used for the left eye. These specific landmark IDs correspond to the extreme positions of the eyes. Based on the selected landmarks, rectangles are created to define the area of each eye. These rectangles serve as a reference frame for cropping the image to include only the area around each eye.



Fig. 1. An example facial landmarks (left) and a clipping for the right eye (right).



Fig. 2. Ptosis detection process.

3.3. Filter Application and Information Extraction

This step involves a series of image transformations designed to enhance key features that are essential for detecting changes related to ptosis. Each transformation step has its own specific task and contributes to better isolation of the anatomical structures.

Image Inversion: The first step is to invert the image color. The purpose of this operation is to increase the contrast between the eyelid and the eye, which facilitates further analysis.

Conversion to Gray Scale: The image is then converted to grayscale. This step reduces the complexity of the image by removing color information and focusing only on the light intensity. This makes it easier to identify edges and other important features of the image.

Image Erosion: Erosion helps remove fine white noise from the image and separate objects in the image that are close together.

Image Binarization: The last transformation is image binarization, which involves applying a threshold that transforms the image into binary (black and white). All pixels with a value above the set threshold become white and the remaining pixels become black. This operation allows for even better separation of the eye area from the rest of the image.

3.4. Pupil Search

Binary image processing is used to detect the location of the iris and pupil.

Contour Search: First, the contours are extracted from the binarized image. We are able to find boundaries between different areas of the image, in this case between the iris and the rest of the eye.

Selection of the Largest Contour: Of the contours found, the one with the largest area is selected. The largest contour is assumed to correspond to the iris, which is crucial for further analysis.

Calculation of the Minimum Surrounding Circle: The minimum circle surrounding the selected contour is then calculated. The center of this circle is taken as the position of the pupil, and its radius is taken as the size of the iris.

Returning Center and Radius: The coordinates of the center of the found circle and its radius are returned. These are key data needed to assess the condition of the eye and the possible appearance of ptosis.

3.5. Eyelid Search

The next step is to extract the contour of the eyelid and evaluate its characteristics.

Determining the Eyelid Contour: First, a binary image is processed to isolate the lines and shapes that correspond to the edges of the eyelid. This operation involves iteratively going through the image columns and determining the points that represent the upper edge of the eyelid. *Image Component Labeling:* Next, the Connected Component Labeling technique is applied to identify the different parts of the image and select the one most likely to correspond to the eyelid.

Parabola to Contour Matching: Once the contour of the eyelid has been isolated, an attempt is made to match the parabolic shape to the designated contour. The parabola is chosen for its ability to accurately replicate the natural shape of the eyelid.

Alternative - Line Matching: When the coefficient a in the parabola equation is negative (which may indicate that the parabola cannot be matched), line matching is used.

3.6. Ptosis Decision

Finally, the algorithm has to decide whether the observed changes in the position of the eyelid relative to the pupil indicate the presence of ptosis. This decision is based on the comparison of the position of the eyelid with that of the pupil.

Comparison of the position of the eyelid and the pupil: The key criterion in the evaluation is whether the eyelid is above or below the pupil. If the eyelid is above the pupil, it is considered that ptosis is not present. If the eyelid is below the pupil, ptosis is diagnosed.

Use of the Eyelid Matching Function: On the basis of a previously matched function (parabola or line), the position of the eyelid in relation to the pupil is determined.

Calculation and Decision: On the basis of the coordinates of the pupil and the equation of the eyelid function, a calculation is made to determine whether the eyelid line intersects the level of the pupil. If so, the diagnosis of ptosis is made.

3.7. Sequential Analysis in Ptosis Detection

After detecting ptosis in individual frames, Predictor proceeds to the sequential analysis stage to assess the continuity of ptosis occurrence and avoid errors caused by accidental blinks.

Grouping of Frames: To reduce the impact of accidental blinks, the predictor analyzes groups of 10 consecutive frames. Within each group, it assesses whether the majority of them show the presence of ptosis.

Decision on the Occurrence of Ptosis in a Group: If a majority of the 10 frames exhibit ptosis, the entire group is considered to have ptosis. Otherwise, it is assumed that ptosis did not occur in this group of frames.

Visualization of Results in Graph: Based on the sequential analysis, a graph is created that shows the moments of occurrence of ptosis for both eyes. This graph is a graphical representation of the continuity of ptosis occurrence throughout the study.

Awarding of Scores for the Study Phase: The final score for this phase of the study is determined by the moment when the ptosis occurs continuously until the end of the study. If ptosis begins at a given time point

and does not cease until the end of the study, points are awarded from that point on.

4. Conclusions

The system developed, including the automated ptosis detection, was practically verified in the Department of Neurology and Vascular Diseases of the Nerve System of the Poznan University of Medical Sciences. Preliminary tests conducted show the usefulness of the method in diagnosing symptoms of ptosis. However, it should be emphasized that due to the small number of patients, we do not yet have a sufficiently large sample at this stage of practical use. Nevertheless, in the case of patients tested in the clinic, the results of the system's performance were consistent with physician observation. It shows sufficient precision to monitor disease status.

References

[1]. R. T. Rousseff, Diagnosis of myasthenia gravis, *Nat. Rev. Dis. Primers*, Vol. 5, Issue 1, 2019, 30.

- [2]. S. Baliński, P. Śniatała, J. Weissenberg, M. Fechner, L. Rzepiński, S. Michalak, Myasthenia gravis disease diagnosis system, in Proceedings of the 31st International Conference on Mixed Design of Integrated Circuits and System (MIXDES'24), 2024, pp. 299-304.
- [3]. Y. Nam, T. Song, J. Lee, et al., Development of a neural network-based automated eyelid measurement system, *Sci. Rep.*, Vol. 14, 2024, 1202.
- [4]. M. Lootus, L. Beatson, L. Atwood, et al., Development and assessment of an artificial intelligence-based tool for ptosis measurement in adult myasthenia gravis patients using selfie video clips recorded on smartphones, *Digit Biomark.*, Vol. 7, Issue 1, 2023, pp. 63-73.
- [5]. J. Kyun Oh, R. Shinder, N. M. Hodgson, Ptosis with fluctuating diplopia, *Archives of Ophthalmology* (1960), Vol. 140, Issue 5, 2022, 538.
- [6]. K. Patel, S. Carballo, L. Thompson, Ptosis, *Dis. Mon.*, Vol. 63, Issue 3, 2017, pp. 74-79.
- [7]. A. Krolak, Vision-Based Eye-Blink Detection System for Mental Fatigue Monitoring and Human-Computer Interfacing, https://repozytorium.p.lodz.pl/bitstreams/4b98a0cca6ce-4bb5-8fd2-b4842b597db2/download

(035)

Identification of Musical Instruments in Audios using Signal Analysis and Artificial Intelligence

A. S. Vazquez-Robledo¹, R. A. Lizarraga-Morales², and M. Lopez-Ramirez¹

¹University of Guanajuato, DICIS, Department of Multidisciplinary Studies, Col. Yacatitas,

Yuriria, Gto, Mexico

² University of Guanajuato, DICIS, Art and Business Department,

Salamanca Valle de Santiago Highway Km. 3.5 + 1.8, Palo Blanco, Salamanca, Gto., Mexico Tel.: + 52 4381144271

E-mail: as.vazquezrobledo@ugto.mx, ra.lizarragamorales@ugto.mx, lopez.misael@ugto.mx

Summary: Music plays an important role in the development of each person's cultural identity, helping to express emotions, telling stories and creating connections among individuals and communities. Music Information Retrieval (MIR) is an emerging field based on software systems designed to extract and retrieve information from music audio files. Some of its main tasks allow automatic analysis of audio signals and extract relevant information, such as the genre, artist, mood, or musical instruments. In this paper, the automatic recognition of 31 musical instruments is proposed. In our proposal, we firstly extract Mel Frequency Cepstral Coefficients (MFCC) and use them as input in an Artificial Neural Network-based classifier. Results show that our proposal is competitive, obtaining results of 97.5 % accuracy for 20 classes and 96.4 % accuracy for 31 classes of musical instruments from a standard dataset.

Keywords: Musical instrument identification, Mel frequency cepstral coefficients (MFCCs), Machine learning, Multi-layer perceptron (MLP), Support vector machines (SVM), Nearest neighbors (KNN).

1. Introduction

Music is a universal element that is part of the daily life of millions of people. Its analysis through signal processing makes it possible to extract relevant information, such as instrument identification or musical structure. Recently, professionals in the field of digital music management have faced a great challenge due to the growth of available data and because of the complexity of data organization. That is why one of the main functions of Music Information Retrieval (MIR) systems is to automatically analyze musical pieces and extract necessary information in order to manage such musical pieces [1]. Some of the main tasks of MIR systems focus on functions that extract artist identification, genre classification, mood classification, musical notation, and the identification of musical instruments. This is crucial for several tasks such as retrieval, sound-source separation, and automatic music transcription.

In the area of musical instrument identification, we can find different techniques from the field of Machine Learning (ML), for example, the work of S. Prabavathy [2]. Prabavathy proposes the automatic classification of musical instruments such as trombone, tuba, trumpet and piano using SVM and the K-Nearest Neighbor (KNN) technique. As part of the results, the manuscript shows an accuracy with SVM of 99.37 % using these techniques. Mahanta et al. [3] presents an Artificial Neural Network (ANN) trained to perform classification of 20 different classes of musical instruments of the London Philharmonic Orchestra in conjunction with the extraction of Mel Frequency Cepstral Coefficients (MFCC), in this work an accuracy of 97 % was achieved. The main advantage

of using Machine Learning in Musical Instrument Identification in Audio Signals is the ability to identify complicated patterns that may be difficult to detect using other techniques, a disadvantage is the high dependency on training data, which must be extensive and with viable features for classification.

Recently, one of the most popular methodologies are those based on Deep Learning. Among the options, one relevant technique is the application of Convolutional Neural Networks (CNN or ConvNets). These refer to a specialized type of Artificial Neural Network specifically designed to process data such as images or audio signals, spectrograms are used as inputs to CNN to learn patterns of how different musical instruments are displayed. Maciej Blaszke [4] introduces the construction of an algorithm for the automation and identification of instruments present in an audio extract, using sets of individual CNN per instrument. The instruments are bass, drums, guitar and piano. In this work, the model efficiency is high, with the metric values ranging from 0.86 for the guitar to 0.99 for drums. A similar architecture is VGGNet, also known as Visual Geometry Group Network, is a CNN architecture.

Chinmay Relkar [5] presents a 4-layer CNN, ConvNet inspired by AlexNet, which is named VGGNet, in this work, we present a score of the evaluation metric F1 of 0.631 (micro) and 0.539 (macro) in the task of instrument recognition in polyphonic music. Relkar also mentions the Regionbased Convolutional Neural Network (RCNN) technique, this technique was one of the first architectures to address the problem of object detection in images using CNN. Although, deep learning-based techniques achieve interesting results, one of the main disadvantages is that these approaches need large amounts of data to generalize well. Whereas machine learning can work with smaller data sets. In addition, deep learning requires a lot of tuning and optimization of hyperparameters to obtain good results.

In this paper, we propose an approach that combines the extraction of Mel Frequency Cepstral Coefficients (MFCCs) and a comparison of different ML approaches such as Multi-Layer Perceptron (MLP), Support Vector Machines (SVM) and Nearest Neighbours (KNN), for 20 and 31 classes of musical instruments. The instruments we explore are: Acoustic Guitar, Alto Saxophone, Balalaika, Bright Piano, Cello, Clarinet, Bowed Double Bass, Pizzicato Double Bass, Drums, Electric Bass, Clean Electric Guitar, Crunch Electric Guitar, Solo Electric Guitar, Electric Piano, Erhu, Flugelhorn, Flute, Fujara, Jinghu, Morin Khuur, Bass Organ, Pan Flute, Piano, Shakuhachi, Sitar, Tenor Saxophone, Trombone, Trumpet, Ukelele, Viola, Violin. We use audio files from the Artificial Audio Multitracks Dataset (AAM) introduced by Ostermann et al [6]. The main proposal is the use of simple approaches that do not require special computational power for the recognition of musical instruments.

2. Methodology

The methodology is presented in Fig. 1, where we can observe the two main phases are proposed: training and evaluation, both phases consist of 3 important parts: Preprocessing of audio files, Extraction of MFCCs features and the use of Machine Learning (ML) techniques. In this proposal, we explore Multilayer perceptron (MLP), Support Vector Machines (SVM) and K-Nearest Neighbors (KNN). In the following sections, each of the parts will be described in more detail.



Fig. 1. Proposed methodology.

3.1. Preprocessing of Audio Files

The analysis of the audio signals begins with the use of the Artificial Audio Multitrack (AAM) dataset introduced by Ostermann et al [6]. For this work, a total of 9300 audio files were taken: 300 audio files for each of the 31 classes of musical instruments contained in this dataset, making it a balanced selection of file numbers for each class. It is worth mentioning that the audio files were preprocessed by eliminating the initial and final silences.

3.2. Feature Extraction

In this paper, we propose to extract 13 of the Mel Frequency Cepstral Coefficients (MFCC) from each of the digital audios. The MFCC model the way humans perceive sound, providing a compact and robust representation of the signal of each audio we will analyze. They capture the most relevant spectral characteristics of each audio and are very useful for audio signal classification, since they represent both the envelope of the spectrum and its changes over time. Providing a compact and robust representation of the signal of each audio [7]. The computation of the MFCC is described as follows.

In order to obtain the MFCCs, the following steps are necessary:

Pre-emphasis: A pre-emphasis filter is applied to increase the energy of the high frequencies and reduce the DC offset (See Eq. (1)).

$$H(z) = 1 - az^{(-1)} 0.9 < a < 1,$$
(1)

where *a* is typically 0.95. H(z) is the filter in the frequency domain, 1 refers to passing the current sample as is (gain=1), $az^{(-1)}$ refers to the output is equal to the current signal minus a fraction of the previous signal. *a* refers to the filter strength (how much emphasis is implemented at high frequencies between 0.9 and 1).

Framing: The signal of each audio file we take is divided into short blocks called frames. The typical frame length is 20-30 ms (milliseconds) and the offset is 10 ms (milliseconds).

Windowing: display of information in a window or frame, where each frame is multiplied by a window using the mathematical function for smoothing the edges of a segment called Hamming, to reduce discontinuities: where N is the length of the frame, the formula is presented below. (See Eq. (3)).

$$h(n) = x(n)w(n), \tag{2}$$

$$w(n) = 0.54 - 0.46 \cos\left(\frac{2\pi}{N-1}\right),$$
 (3)

where w(n) is the value of the window in the sample n, where n ranges from 0 to N - 1, 0.54 refers to the constant component (DC term of the window). $-0.46 \cos\left(\frac{2\pi}{N-1}\right)$ refers to the oscillatory component that gives the window a smooth shape.

Spectral estimation: for this step the Discrete Fourier Transform (DFT) is applied to each frame to obtain the spectral coefficients, the formula is presented below (See Eq. (4)).

$$x(k) = \sum_{n=0}^{N-1} y(n) \cdot e^{-j\frac{2\pi}{N}kn},$$
 (4)

$$0 \le n, \ k \ge N - 1,\tag{5}$$

where x(k) refers to the transformed value at frequency k. k can also be said to be the result of the Discrete Fourier Transform at index k, N is the total number of samples of the signal.

The frequency scale of each audio is transformed to the Mel scale, this scale focuses on how humans perceive sound frequencies, which is closer to human perception (See Eq. (6)).

$$f_M = 2525 \times log\left(1 + \frac{f}{700}\right),$$
 (6)

where f_M is the Mel frequency for the linear frequency f. Once we have the Mel scale, the logarithm of the energy of the magnitude of the Mel filter response is taken as shown in Equation (7).

$$E_{j}^{x} = \sum_{n=1}^{n} |x(k)|^{2} * \psi_{i}(k), \qquad (7)$$

where |x(k)| is the amplitude spectrum, k is the frequency index, ψ_i are the i^{th} Mel band pass filter, $1 \le i \le M$, and M is the number of Mel-scaled triangular band-pass filters. E_j^x is the filter bank energy. Finally, the Discrete Cosine Transform (DCT) is applied to the logarithms of the energy to obtain the Cepstral Coefficients of each audio (See Eq. (8)).

$$C_t^x = \sum_{t=1}^M \log(E_t^x) Cos\left[l \cdot \frac{(2\pi - 1)\pi}{2M}\right],\tag{8}$$

where C_t^x describes the calculation of a cepstral coefficient part of the MFCCs process, *M* is the total number of frequency bands. $\log(E_t^x)$ is the logarithm of the spectral energy in the band *t* This is typical in the extraction of cepstral coefficients or MFCCs, since applying logarithm compresses the signal dynamics. $\left[l \cdot \frac{(2\pi-1)\pi}{2M}\right]$ refers to the cosine-weighted term.

3.3. Classification Techniques

The first 13 MFCCs were used as input features to classification systems. In this manuscript we explore the performance of 3 different classic machine learning approaches: a multilayer perceptron (MLP), Support Vector Machines (SVM) and Nearest Neighbors (KNN). We use 80 % of the files from the dataset are used for training and the remaining 20 % are used for evaluation.

The architecture for the MLP consists of an input layer of 13 neurons where the relevant features are selected from the first 13 Mel Frequency Cepstral Coefficients (MFCC) of the audio files. Then, the network consists of 3 hidden layers, the first layer with 256 neurons, the second layer with 128 neurons and the third layer with 256 neurons. These neurons are activated by the ReLU function, which allows the model to learn intermediate representations and complex patterns of data. The output layer uses a SoftMax function to convert the outputs into probabilities consistent with as many neurons as classes to be predicted. The model is trained for 2000 epochs using batches of 32 samples, allowing continuous adjustment of the weights after each batch, optimizing the efficiency of the training process.

The SVM architecture consists of an algorithm that finds the optimal hyperplane that best separates the classes of the musical instruments in a feature space by maximizing the margin between the closest instances of different classes, known as support vectors. The ECOC (Error-Correcting Output Codes) technique, which focuses on binary classifiers for multi-class classification problems, was implemented with the RBF (Radial Basis Function) kernel, which is used to obtain non-linear data separable into a higher dimensional space where the classes of the musical instruments can be linearly separated.

The KNN architecture consists in that when a new unlabelled data is presented, the algorithm compares it with all the data in the training set and selects the k most similar examples, where k is an integer representing the number of neighbours to be considered. In this case K will be equal to 5 neighbours and the Euclidean distance will be implemented to measure similarities and thus consider which class it belongs out of the classes of musical instruments.

3. Experimental Results

In this paper, we evaluate the performance of our proposal by using evaluation metrics such as Precision (P), Recall (R), F1-score(F) and Accuracy (A). Additionally, cross-validation with 5 folds was implemented. In the first experiment, we used only 20 instruments (classes), in order to make a comparison with by Mahata et al [2]. In Table 1, we can observe the results for 20 classes, the MLP managed to outperform the work proposed by Mahata et al [2]. In Accuracy, Mahanta obtains 97 % and the MLP obtains 97.5 %. SVM and KNN obtained 95.2 % and 93.4 %, respectively.

 Table 1. Performance evaluation and comparison 20 classes.

Approach	P (%)	R (%)	F (%)	A (%)
Mahanta [2] (20 classes)	97.0	97.0	97.0	97.0
MLP (20 clasess)	97.5	97.5	97.4	97.5
SVM (20 classes)	95.2	95.5	95.3	95.2
KNN (20 classes)	93.4	93.6	93.4	93.4

In the second experiment, we increased the number of classes to 31 instruments and explored which classification method achieves better results. In Table 2, we can observe the results for 31 classes, the MLP achieved the best results. In Accuracy, the MLP scored 96.4 %. SVM scored 94.0 % and KNN scored 91.7 %. The proposed system was implemented in MATLAB.

Table 2. Performance evaluation and comparison30 classes.

Approach	P (%)	R (%)	F (%)	A (%)
MLP (31 classes)	96.4	96.5	96.4	96.4
SVM (31 classes)	94.0	94.3	94.1	94.0
KNN (31 classes)	91.7	91.9	91.6	91.7

4. Conclusions

Classification of musical instruments was carried out, using Machine Learning techniques and MFCCs. The model shows a classification performance with an accuracy of 96.4 % for 31 classes and 97.5 % for 20 classes in the best case with the use of the Multilayer Perceptron (MLP). The computation of MFCC features has been shown to be simple, yet they are robust enough to describe musical instruments. Experiments on an extensive dataset show that our method yields higher accuracy, outperforming other systems proposed for the same task in the state-of-the-art.

References

- [1]. A. Lucena, C. Pires, K. Nose-Filho, R. Suyama, Musical instruments recognition using machine learning, in *Proceedings of the Brazilian Technology Symposium*, Brazil, Oct. 2020.
- [2]. S. V. R. Prabavathy, Musical instruments classification using pre-trained model, *International Research Journal of Engineering and Technology (IRJET)*, Vol. 7, Issue 5, May 2020, pp. 585-589.
- [3]. S. K. Mahanta, A. F. U. R. Khilji, P. Pakray, Deep neural network for musical instrument recognition using MFCCs, *Computación y Sistemas*, Vol. 25, Issue 2, 2021, pp. 351-360.
- [4]. M. Blaszke, B. Kostek, Musical instrument identification using deep learning approach, *Sensors* (*Basel, Switzerland*), Vol. 22, Issue 8, 2022, 3033.
- [5]. V. T. Chinmay Relkar, Musical instrument identification using, International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET), Vol. 2, Issue 9, September 2019, pp. 1826-1829.
- [6]. F. Ostermann, I. Vatolkin, M. Ebeling, AAM: a dataset of artificial audio multitracks for diverse music information retrieval tasks, *EURASIP Journal on Audio, Speech, and Music Processing*, 2023, 13.
- [7]. S. A. Majeed, H. Husain, S. A. Samad, T. F. Idbeaa, MEL frequency cepstral coefficients (MFCC) feature extraction enhancement in the application of speech recognition: A comparison study, *Journal of Theoretical and Applied Information Technology*, Vol. 79, Issue 1, 2015, pp. 2005-2015.

(038)

Enhancing Real-time Decision-making with Scalable, Safe, and Private LLMOps and Context-aware RAG Workflows

Jérémie Farret, Jerin Jude and Nitish Kumar Pilla

Mind in a Box Inc. 3575 St Laurent Boulevard, Suite 200 Montreal, Quebec H2X 2T6 Canada Tel.: +1-833-636-2269 E-mail: jeremie@mindinabox.ai

Summary: This study explores the integration of scalable Large Language Model Operations (LLMOps) with advanced Retrieval-Augmented Generation (RAG) workflows to address challenges in deploying high-performance AI systems. It emphasizes retrieval's role as a critical component in RAG, ensuring relevance and accuracy. The architecture achieves enhanced relevance, efficiency, and scalability using technologies like vLLM for low-latency inference, Nvidia A30 GPUs for accelerated processing, and OpenSearch for hybrid search. Components like hybrid search, reranking, HyDE, and domain-specific embedding adapters optimize retrieval and generation processes. Kubernetes and Docker facilitate dynamic scaling and resource management, while on-premises deployment prioritizes data privacy. The SciFact dataset is used to evaluate the system's retrieval and generation performance, with metrics like NDCG and MAP assessing effectiveness. The study highlights incremental improvements from enhanced RAG features and test scalability under high query loads, demonstrating a robust, efficient solution for sensitive, high-stakes applications.

Keywords: GPU-accelerated inference, Large Language Models (LLMs), RAG workflows, Benchmarking, Inference efficiency, Kubernetes, Mind in a Box.

1. Introduction

Large Language Models (LLMs) have transformed real-time decision-making across industries. However, the effectiveness of LLMs is inherently limited by the quality of the context provided to them - often stated, "LLMs are only as good as the context they receive." While LLMs demonstrate remarkable generative capabilities, their reliance on static, pre-trained knowledge poses challenges in dynamic, real-world applications where up-to-date and domain-specific information essential. Retrieval-augmented is generation (RAG) addresses this limitation by incorporating real-time retrieval mechanisms, ensuring that responses are grounded in accurate, relevant, and fresh data.

Despite the advantages of RAG, retrieval remains a critical bottleneck. Traditional retrieval systems often struggle with balancing relevance and efficiency, especially for ambiguous or specialized queries. This study focuses on optimizing retrieval in RAG workflows through hybrid search techniques, reranking strategies, and domain-adaptive embedding models. By integrating scalable LLMOps with enhanced retrieval mechanisms, we aim to improve response accuracy, minimize latency, and ensure robust decision-making in sensitive, high-stakes applications.

2. RAG (Retrieval Augmented Generation)

Retrieval-Augmented Generation (RAG) improves AI-generated responses by integrating retrieval-based systems with generative models, ensuring outputs are grounded in factual information. Unlike standalone generative models that rely solely on pre-trained knowledge, RAG retrieves relevant documents from external sources, reducing hallucinations and improving adaptability for knowledge-intensive tasks such as customer support, research assistance, and legal or medical analysis. By dynamically incorporating updated information, RAG allows large language models (LLMs) to generate more precise, context-aware, and up-to-date responses.

However, effective retrieval remains a major challenge, as standard retrieval methods may fail to fetch the most relevant documents, particularly for vague, complex, or highly specialized queries. Sparse retrieval methods, such as keyword-based searches BM25), often struggle with semantic (e.g., understanding, while dense retrieval models using embeddings may overlook exact keyword matches. This imbalance can lead to over-retrieval of loosely related documents or under-retrieval of critical information. Additionally, the quality of retrieved results heavily depends on the structure and completeness of the knowledge base, making retrieval a significant bottleneck in high-stakes applications like legal review, medical diagnosis, and scientific discovery.

To address these challenges, modern RAG workflows employ several advanced retrieval techniques. **Hybrid Search** [3] combines sparse retrieval (like BM25) with dense retrieval (using neural embeddings) to ensure both keyword-specific and semantic relevance. Sparse methods excel at retrieving exact keyword matches, while dense retrieval captures conceptual relationships, improving recall and precision. **Reranking** further refines retrieval results by applying transformer-based models to score and reorder documents, ensuring that only the

most relevant ones are passed to the LLM. Hypothetical Document Embeddings (HyDE) [1, 7] generate synthetic documents based on the user's query, enriching the retrieval space and improving performance in sparse or highly specialized domains. Embedding adapters [2, 8] fine-tune pre-trained embeddings for specific fields like medicine or law, optimizing retrieval without requiring complete retraining. Together, these techniques enhance retrieval quality, making RAG more reliable for complex, domain-specific applications.

3. LLMOps

The infrastructure system selected for the experimental protocol, Mind in a Box (M/B) Catalyst, operates a package called Mind in a Box AI+ which integrates and operationalizes both LLMOps and the vLLM Inference Server with Kubernetes, in a turn-key Equipment as a Service solution. This provides a solid foundation for efficient, scalable, and resilient LLMOps pipelines. Kubernetes provides the orchestration layer, ensuring consistent deployment and management of containerized LLM workflows across diverse infrastructures. Its features, such as auto-scaling, workload balancing, and fault tolerance, enable dynamic scalability and high availability, ensuring smooth operations even during varying query loads. Role-based access control (RBAC) and secure namespaces further enhance data privacy and regulatory compliance, making Kubernetes an component for essential managing sensitive applications.

Within this orchestrated framework, the M/B Catalyst infrastructure and the supported inference server optimize LLM performance by enabling lowlatency and high-throughput inference. By leveraging on one hand advanced memory management techniques such as dynamic caching and efficient tensor partitioning, it maximizes GPU utilization and minimizes computational overhead. On the other hand, a proprietary data bus enables high-performance concurrency, which ensures responsiveness, making it ideal for real-time, high-demand applications. Seamless integration with APIs and compatibility with various LLM architectures further simplify deployment and operation. Together, Kubernetes and the inference server provide a robust and integrated solution for scalable LLMOps. Kubernetes handles deployment, scalability, and fault tolerance, while the inference server focuses on efficient inference, ensuring that LLMs operate at peak performance with minimal latency. Coupled with a high-performance computing solution targeted at reducing concurrency bottlenecks, such as the proprietary M/B Catalyst system, the integrated architecture aims to support peak inference performances for on-premise and hybrid topologies. Not only does this synergy enable organizations to deploy and manage large-scale LLMs in a reliable, efficient, and secure manner, addressing the demands of modern, high-performance AI

applications. But coupled with an efficient data bus infrastructure, it enables to support LLMops with comparatively much smaller energy footprints and waste heat emissions than similar GPUaaS-based LLMops propositions.

4. Experimentation Setup and Methodology

The experimentation setup evaluates a scalable Large Language Model Operations (LLMOps) pipeline integrated with advanced Retrieval-Augmented Generation (RAG) workflows. This section outlines the infrastructure, datasets, evaluation metrics, and methodology employed.

4.1. Infrastructure

The general architecture was supported by an on-premises combination of Mind in a Box Catalyst (GPU-based for LLMOps) and Mind in a Box Zen (CPU-based DataOps) high-performance for computing systems. The hardware infrastructure proposed by the M/B Catalyst for LLMOps includes Nvidia A30 GPUs, which support accelerated tensor computations and optimized inference capabilities, alongside 48-core Intel Xeon Gold 6338N processors for high-performance task execution. Additionally, the Mind in a Box Zen DataOps cluster was equipped with 96 (24x4) GB of RAM and 2 TB of SSD storage to handle data efficiently and ensure effective caching mechanisms.

This infrastructure is used according to two modalities. A first one, purely on premise, as illustrated below in Fig. 1, where the LLMops architecture described in the previous chapters is deployed exclusively on premise.



Fig. 1. LLMops architecture employed for a purely on-premises workflow modality.

A second modality, illustrated below in Fig. 2, is similar but based on a hybrid workflow, using the OpenAI LLM PaaS services and APIs.

The software infrastructure incorporated Kubernetes (v1.25) for managing containerized workflows. Kubernetes was configured with rolebased access control (RBAC) and secure namespaces to enhance data privacy. For inference, the vLLM server was employed to deliver low-latency and high-throughput large language model inference. OpenSearch, deployed on the M/B Zen DataOps system as the hybrid search engine, supporting both vector embeddings and BM25 ranking mechanisms. Python (v3.10), PyTorch (v2.0), and LangChain were utilized for the seamless integration of LLM and retrieval components, while Docker (v24) was used for the packaging and deployment of RAG components and LLM.



Fig. 2. LLMops architecture employed for a hybrid workflow modality.

4.2. Dataset and Models

The study utilized the SciFact [6] dataset from the BEIR benchmark [5], a resource specifically designed for scientific claim verification. The dataset consists of expert-written scientific claims paired with annotated abstracts from scientific literature, which serve as evidence. These abstracts are labeled with veracity and rationales, indicating whether they support or refute the claims. This structure allows for a comprehensive evaluation of a system's ability to retrieve relevant information and assess the validity of scientific statements effectively.

For the models, the study employed a combination of advanced embedding and generative techniques. The embedding model used was BAAI/bge-small-env1.5, developed by the Beijing Academy of Artificial Intelligence. This model transforms input text into 384-dimensional vectors, providing efficient semantic representation and similarity computation while maintaining a balance between performance and computational efficiency. To generate Hypothetical Document Embeddings (HyDE) queries, the study utilized gpt-4o-mini-2024-07-18, a smaller and cost-effective variant of OpenAI's GPT-40 series. Despite its reduced size, this model maintains state-of-the-art intelligence and is well-suited for generating high-quality synthetic queries to enhance retrieval. Additionally, a linear adapter module was implemented to fine-tune the embeddings for the specific task. This adapter, consisting of a single linear layer, refines the embeddings using a triplet margin loss function with a margin parameter of 1.0, optimizing their ability to distinguish between relevant and non-relevant documents for the verification task.

4.3. Evaluation Metrics

The system's retrieval performance was evaluated using key metrics at cutoff points of 2, 5, and 10 to ensure high-quality and contextually relevant results. Normalized Discounted Cumulative Gain (NDCG) assessed the ranking of results, prioritizing highly relevant documents early in the list. Mean Average Precision (MAP) measured precision across recall levels, ensuring a balance between completeness and relevance. Additionally, Recall evaluated the system's ability to retrieve all relevant documents, while Hit Rate measured the likelihood of at least one relevant document appearing in the top-k results. These metrics ensured the system delivered comprehensive and accurate outputs, critical for tasks relying on precise retrieval.

Mean Reciprocal Rank (MRR) was also included to evaluate how quickly the first relevant document was retrieved, minimizing delays in accessing key information. Together, these metrics provided a robust assessment of the system's ability to prioritize, retrieve, and present relevant results effectively. This evaluation ensured the system met the high demands of tasks where the quality, ranking, and speed of retrieved information directly impact performance.

4.4. Methodology

This study benchmarked and compared retrieval techniques to enhance **Retrieval-Augmented** Generation (RAG) workflows. The methodology included data ingestion, baseline retrieval evaluation, advanced retrieval implementation, and systematic evaluation. A scientific corpus was preprocessed into structured LangChain Document objects for compatibility and experimentation. Dense retrieval used the BAAI/bge-small-en-v1.5 embedding model to encode documents into vector representations, indexed in an OpenSearch vector database with SSL, authentication, and bulk ingestion for security and scalability. This enabled semantic retrieval even without explicit lexical matches. A BM25Retriever was also trained on the corpus for sparse retrieval using TF-IDF scoring, stored as a pickle file for consistent benchmarking.

The retrieval pipeline began with BM25-based sparse retrieval for a baseline evaluation, followed by dense retrieval using BAAI/bge-small-en-v1.5 embeddings. A hybrid strategy combined BM25 and dense retrieval scores to improve precision and recall.

The FlashRank reranker refined the top 50 documents, reordering them for contextual relevance, with the top 10 selected for evaluation. Hypothetical Document Embeddings (HyDE) were introduced to address claim ambiguities, using the gpt-40-mini-2024-07-18 model to generate synthetic passages as augmented queries, bridging contextual gaps. Finally, a task-specific linear adapter was fine-tuned on the SciFact dataset using triplet margin loss, aligning query embeddings with positive documents and distancing them from negatives, with dynamic random negative sampling for robustness.

Metric	Cut off	BM25	Dense	Hybrid	Re Rank	HyDE	Adapter
	@2	0.53	0.65	0.65	0.67	0.68	0.76
Recall	@5	0.63	0.77	0.76	0.79	0.83	0.81
	@10	0.68	0.87	0.84	0.87	0.91	0.84
	@2	0.55	0.68	0.68	0.70	0.72	0.78
Hit Rate	@5	0.65	0.79	0.78	0.82	0.86	0.83
	@10	0.71	0.88	0.86	0.88	0.92	0.85
	@2	0.51	0.64	0.63	0.64	0.67	0.75
nDCG	@5	0.55	0.69	0.68	0.70	0.74	0.77
	@10	0.57	0.72	0.71	0.72	0.77	0.78
	@2	0.50	0.61	0.61	0.61	0.65	0.73
MAP	@5	0.52	0.65	0.65	0.65	0.71	0.75
	@10	0.53	0.67	0.66	0.67	0.72	0.75
	@2	0.52	0.64	0.64	0.63	0.68	0.75
MRR	@5	0.54	0.67	0.67	0.67	0.72	0.76
	@10	0.55	0.68	0.68	0.68	0.73	0.76

Table 1. Performance comparison of various retrieval methods.

The loss curve, shown below, illustrates the steady convergence of the model during training (Fig. 3).



Fig. 3. Training Loss Curve.

The training process was implemented using PyTorch, with a lightweight linear adapter layer. Optimization was performed using the AdamW optimizer with a learning rate of 0.003, a linear warmup scheduler, and gradient clipping to stabilize training. The model was trained for 50 epochs with a batch size of 32, during which the triplet margin loss consistently decreased, indicating improved embedding alignment for query-document relevance. This fine-tuning step enhanced the dense retrieval pipeline's ability to deliver task-specific relevance for scientific fact-checking workflows.

The system's retrieval performance under highdemand scenarios was evaluated using a structured query load testing methodology. The ShareGPT and dataset. comprising diverse realistic conversational prompts, was used to simulate real-world usage. Queries were preprocessed to ensure appropriate token lengths, maintaining the representativeness of typical usage scenarios. The neuralmagic/Meta-Llama-3-1-8B-Instruct-FP8 model, optimized for high-performance inference, was used for benchmarking. It was configured with a maximum context length of 16,384 tokens to handle complex conversational tasks. The serving framework,

supported by Kubernetes, enabled dynamic batching, efficient token scheduling, and resource allocation to manage concurrent requests with minimal latency.

Testing involved dispatching all queries simultaneously to simulate extreme burst load scenarios, mimicking sudden spikes in demand. Key performance metrics, including retrieval latency, query throughput, and resource utilization, were monitored. Additional granularity was achieved by analyzing time to first token (TTFT), time per output token (TPOT), and inter-token latency (ITL). This comprehensive evaluation assessed the system's scalability, responsiveness, and resource efficiency under extreme stress conditions.

5. Results

The following section discusses the retrieval accuracy metrics and performance metrics results obtained after experiments.

Benchmarking results highlight performance differences among retrieval methods - BM25, Dense, Hybrid, Reranking, HyDE, and Adapter - across metrics like recall, hit rate, MRR, MAP, nDCG, and R-Precision. BM25, a traditional method, shows the weakest performance (recall@10: 0.68, hit rate@10: 0.71), serving as the baseline. Dense retrieval improves significantly (recall@10: 0.87, hit rate@10: 0.88), demonstrating the effectiveness of dense embeddings. Hybrid methods (recall@10: 0.84, hit rate@10: 0.86) perform comparably to Dense, while Reranking slightly enhances results (recall@10: 0.87, hit rate@10: 0.88). HyDE excels with the highest recall@10 (0.91), hit rate@10 (0.92), and nDCG@10 (0.77), showcasing superior relevance and ranking. Adapter leads in early precision (recall@2: 0.76, hit rate(a)2: 0.78) and achieves the highest MRR (0.76) and MAP (0.75), making it ideal for tasks prioritizing top-ranked results. Overall, advanced methods outperform BM25, with the choice depending on specific retrieval goals: Adapter for early precision, HyDE for recall and hit rate.

7th International Conference on Advances in Signal Processing and Artificial Intelligence (ASPAI' 2025), 8-10 April 2025, Innsbruck, Austria



Fig. 4. Comparison of metrics across different retrieval strategies.

Table 2. Throughput Metrics.

Throughput	Value
Request Throughput (req/s)	7.25
Input Token Throughput (tok/s)	1684.97
Output Token Throughput (tok/s)	1409.03

Benchmarking of the on-premises neuralmagic/Meta-Llama-3.1-8B-Instruct-FP8 model shows high efficiency, with a throughput of 7.25 requests/second, input token processing at 1684.97 tokens/second, and output generation at 1409.03 tokens/second. Latency metrics reveal a mean Time to First Token (TTFT) of 46.83 seconds (median: 41.14 s, P99: 104.92 s), indicating occasional delays for complex queries. Token generation speeds are reasonable (mean TPOT: 172.00 ms, median: 150.75 ms), though P99 TPOT spikes to 872.46 ms. Inter-Token Latency (ITL) is efficient (median: 85.92 ms) but peaks at 724.95 ms (P99). The deployment demonstrates strong throughput and scalability for batch tasks but requires latency optimization for real-time applications.

Metric	Mean (s)	Median (s)	P99 (s)
Time to First Token (TTFT)	46.83	41.14	104.92
Time per Output Token (TPOT)	0.172	0.151	0.872
Inter-Token Latency (ITL)	0.372	0.086	0.725

Table 3. Latency Metrics.

6. Conclusion

This study successfully integrates scalable LLMOps with advanced RAG workflows, to enhance relevance, efficiency, and scalability. Techniques such as hybrid search, reranking, HyDE, and embedding adapters significantly improve retrieval and generation performance, as demonstrated by benchmarking on the SciFact dataset using metrics like nDCG, MAP, MRR, HitRate and Recall. The system excels in high-stakes, sensitive applications, offering robust, real-time decision-making capabilities while prioritizing data privacy and security.

References

[1]. C.-M. Chan, C. Xu, R. Yuan, H. Luo, W. Xue, Y. Guo, J. Fu, RQ-RAG: learning to refine queries for retrieval augmented generation, *arXiv preprint*, 2024, arXiv:2404.00610.

- [2]. T. Schopf, D. N. Schneider, F. Matthes, Efficient domain adaptation of sentence embeddings using adapters, in *Proceedings of the 14th International Conference on Recent Advances in Natural Language Processing (RANLP'23)*, Varna, Bulgaria, September 2023, pp. 1046-1053.
- [3]. A. Bansal, Optimizing RAG with hybrid search and contextual chunking, *Journal of Engineering and Applied Sciences Technology*, Vol. 5, Issue 4, 2023, pp. 1-5.
- [4]. W. Kwon, Z. Li, S. Zhuang, Y. Sheng, L. Zheng, C. H. Yu, J. E. Gonzalez, H. Zhang, I. Stoica, Efficient memory management for large language model serving with paged attention, *arXiv preprint*, 2023, arXiv:2309.06180.
- [5]. N. Thakur, N. Reimers, A. Rücklé, A. Srivastava, I. Gurevych, BEIR: a heterogeneous benchmark for zero-shot evaluation of information retrieval models, *arXiv preprint*, 2021, arXiv:2104.08663.
- [6]. D. Wadden, S. Lin, K. Lo, L. L. Wang, M. van Zuylen, A. Cohan, H. Hajishirzi, Fact or fiction: verifying scientific claims, *arXiv preprint*, 2020, arXiv:2004.14974.
- [7]. L. Gao, X. Ma, J. Lin, J. Callan, Precise zero-shot dense retrieval without relevance labels, *arXiv preprint*, 2022, arXiv:2212.10496.
- [8]. J. Yoon, Y. Chen, S. Arik, T. Pfister, Search-adaptor: embedding customization for information retrieval, in *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics*, Vol. 1, 2024, pp. 12230-12247.

(039)

Self-adaptive and Self-learning Lighting System: Integrating LSTM and RL for Energy Efficiency and Personalized Visual Comfort

G. Potenza¹, Cristina Baglivo², M. Bonomolo³ and <u>P. Ribino¹</u>

¹ National Research Council (CNR), Institute for High-Performance Computing and Networking (ICAR), Italy ² Department of Engineering for Innovation, University of Salento, Italy ³ Department of Engineering, University of Palermo, Italy E-mail: patrizia.ribino@icar.cnr.it

Summary: Optimizing lighting systems is crucial for reducing energy consumption and enhancing occupant well-being in sustainable building design. A key challenge is creating energy-efficient lighting systems that adapt to individual users' visual comfort needs. This paper proposes a two-phase approach for a self-adaptive and self-learning lighting control system. In the first phase, Long Short-Term Memory (LSTM) networks optimize the placement of photosensors by modelling dynamic lighting conditions over time. In the second phase, Reinforcement Learning (RL) enables real-time adaptation of lighting based on occupant preferences, maximizing energy efficiency and visual comfort. This system ensures personalized, efficient lighting in office environments while minimizing energy waste.

Keywords: Long short-term memory, Reinforcement learning, Visual comfort, Smart lighting systems.

1. Introduction

Lighting is a major source of energy demand [1], and optimizing lighting systems in sustainable building design is crucial to reducing energy consumption [2]. Moreover, as lighting conditions significantly impact human health and well-being and affect task performance [3], visual comfort is the counterpart to be considered. As energy efficiency becomes increasingly important in environmental sustainability, a key challenge in built environments, particularly in office spaces, is creating energy-efficient lighting systems adaptable to individual users' varying visual comfort needs [4-7].

To address this issue, recent advancements in technologies such as distributed sensing and machine learning have opened new avenues for developing intelligent lighting systems that can dynamically respond to energy efficiency and personalized comfort requirements. However, a critical issue of such systems is the positioning of the photosensors since they monitor ambient light levels and inform lighting control systems to adjust illumination accordingly.

Due to the dynamic nature of lighting conditions – impacted by factors such as time of day, building layout, and individual preferences – if placed optimally, photosensors can reduce artificial lighting in the presence of sufficient daylight, thus reducing overall energy consumption. Indeed, lighting sensors cannot be placed on the work plane because, in this position, they do not effectively capture the full range of environmental lighting conditions and can be obstructed by objects or human movement. It is more effective to place sensors at elevated positions, such as the ceiling or high on walls. Moreover, finding an optimal placement of these sensors also means ensuring a great correlation between the level of illuminance on the work plane and those acquired by the sensors. Hence, the proposed solution introduces a two-phase approach to build a self-adaptive and self-learning lighting control system for energy efficiency and personalized comfort requirements.

In the first phase, photosensor placement optimization is achieved using Long Short-Term Memory (LSTM) networks, which capture temporal dependencies in sequential data. Thus, LSTMs model lighting conditions' dynamic and time-varying nature. Using LSTM models, optimal lighting conditions can be predicted based on historical data. The optimally performed LSTM model indicates the sensor that best supports energy-efficient lighting controls.

The second phase introduces Reinforcement Learning (RL) to enable real-time adaptation of lighting systems. RL allows the system to continuously learn from its interactions with the environment, adapting visual comfort parameters based on occupant preferences. The system dynamically adjusts lighting and illuminance uniformity using a reward-based framework to maximize energy efficiency and visual comfort. This real-time learning ensures that the lighting control system remains flexible and personalized, providing an optimal lighting experience in a workplace environment. This adaptive control system not only enhances occupant comfort but also ensures that lighting energy consumption is minimized without compromising visual quality.

In this paper we present a preliminary case study to assess the strengths of the proposed solution.

2. Methods

2.1. Visual Comfort

Visual comfort is typically assessed by evaluating factors such as the amount of light, light uniformity,

colour rendering quality, and the risk of glare. In this preliminary paper, we consider only the contribution of the amount of light and its uniformity.

Good visibility is defined by adequate amount of light, allowing occupants to accomplish their tasks. Discomfort can be caused by either too low or too high light levels. It is assessed by the illuminance (Eq. (1)):

$$E[lux] = \frac{\Phi[lm]}{A[m^2]},\tag{1}$$

where A is the work plane surface and Φ is the luminous flux on the surface. In typical offices, EN 12464-1 standard [8] suggests a target value of 500 lux.

On the other hand, the illuminance Uniformity (UO) [8] describes how evenly light spreads over a task area. A well-designed uniformity of lighting (UO) helps prevent visual stress by minimizing the need for frequent eye adjustments between over-lit and under-lit areas, thereby reducing the risk of visual discomfort. The Eqs. (2) and (3) allow to compute uniformity:

$$UO_{max} = \frac{E_{min}}{E_{max}},\tag{2}$$

$$UO_{average} = \frac{E_{min}}{E_{average}},$$
(3)

where E_{min} , E_{max} and $E_{average}$ are the work plane's min, max, and average illuminance. Many lighting standards [9] require an $UO_{average} = 0.8$ or $UO_{max} = 0.7$.

Finally, a further element to be considered is the Daylight Factor (DF) [9] that is a measure used to evaluate the amount of natural light entering in a building. It compares the light level inside a space to the light level outside (on an overcast day), providing an indicator of how much natural daylight is available indoors. It is defined as:

$$DF = \left(\frac{E_{in}}{E_{out}}\right) * 100 \%, \tag{4}$$

where E_{in} is the illuminance due to daylight at a point on the indoor working plane, E_{out} is the external horizontal illuminance.

2.2. LSTM

The architecture of the adopted deep neural network is shown in Fig. 1, where the LSTM cells are used as basic building blocks in the hidden layers (see Fig. 2). The input layer mainly processes the data, receiving temporal data organized in time windows. In our model, the inputs are the current illuminance values of a given photosensor, solar elevation and azimuth values, and the illuminance value on the work plane at previous time steps. LSTM layers store long-term information due to their gating mechanisms. Dense layers process the output of the LSTM layers and provide the final prediction, such as the future illumination level on the work plane. Fig. 2 shows the architectural scheme of the LSTM cells. The Root Mean Squared Error is used to evaluate the optimal sensor position; it is calculated as the mean of the squares of the differences between predictions \hat{y} and actual values y of the illuminance on the work plane.



Fig. 1. LSTM Architecture.



Fig. 2. LSTM Cell.

2.3. Reinforcement Learning Model

In the RL model, the agent learns the optimal policy through trials during the interactions with the environment. This interaction process is formulated as a Markov Decision Process, and we use Q-learning algorithms founded on the Bellman Equation:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha \times \\ \times \left[R_t + \gamma \times \max_{a_{t+1}} Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t) \right]$$
(5)

The reward function is formalized according to the following equation:

$$R_{t} = w_{U} \left(1 - \left| \frac{U_{t} - U_{target}}{U_{target}} \right| \right) + w_{E} \left(1 - \left| \frac{E_{t} - E_{target}}{E_{target}} \right| \right)$$
(6)

where U_t and E_t are the values of **uniformity and illuminance** reached at time t, U_{target} and E_{target} are the values of uniformity and illuminance that the agent aims to achieve, as specified by the user. w_U and w_E are weights associated with uniformity and illuminance.

The agent is incentivized to adjust its actions to bring uniformity and illuminance as close as possible to the user-defined target values, and the reward reflects how well it achieves these targets.

3. Case Study

For the experimental setup, we considered a typical office with several photosensors collecting different light exposure levels. One sensor is positioned on the work plane to capture the real illuminance levels, and the others are located at different office points. After model training and the assessment of the optimal sensor, the system is tested at runtime to evaluate energy performance and visual comfort derived from the system's adaptation to user preferences.

In this preliminary assessment, to evaluate visual comfort in a controlled environment, we designed an experimental setup considering a work plane of dimensions 2 m² (i.e., width=2m and depth=1m). The main light source is a desk lamp initially placed at position x=170, y=30, z=50 (with the axes origin placed at the top left corner of the desk), as shown in Fig. 3. Th desk lamp can be moved on the work plane and adjusted in height up to 1.5 meters.



Fig. 3. Reference system with respect to the desk.

The system is configured to assess user preferences based on two primary metrics: illuminance and uniformity. The target illuminance is set between 500 to 700 lux, representing the ideal lighting range for visual comfort in typical work environments. This range accounts for both the adequacy and comfort of light levels for prolonged tasks.

Additionally, the uniformity of illuminance across the workplace is set as a critical parameter, with a target uniformity value of ≥ 0.7 . This ensures a consistent light distribution across the workplace to avoid areas of excessive brightness or insufficient illumination.

Finally, we also consider the relationship between artificial and natural lighting to achieve the optimal balance of illuminance and uniformity, both of which are essential for maintaining visual comfort. Hence, the contribution of natural light is considered by assuming a daylight factor of 2. Although, the outdoor illuminance levels vary throughout the day due to the changing position of the sun and atmospheric conditions and the exact values can shift depending on the weather or other factors, there are general patterns to expect at different times of the day. For this simplified experimental setup, we considered the outdoor illuminance E_{out} for four parts of the day as follows:

- 1. Morning (Post-Sunrise to Noon) E_{out}=10000 lx;
- 2. Noon E_{out}=50000 lx;
- 3. Afternoon (Post-Noon to Sunset) E_{out}=5000 lx;
- 4. Evening (Sunset to Night) E_{out}=500 lx.

These values are used to calculate the contribution of natural light to the overall lighting conditions, working in combination with the artificial light provided by the desk lamp.

3.1. Preliminary Results

To demonstrate the efficacy of the proposed approach, we present preliminary results focused both on energy efficiency and on the achievement of user preferences. compares energy waste using the optimal sensor detected by our approach and a test sensor placed according to typical guidelines concerning the real illuminance level on the work plane. The graphs in Fig. 4 indicate that the test sensor leads to inefficiencies in the lighting control system, causing periods of excessive energy waste. In contrast, the lighting system with the optimal sensor performs much more efficiently, with minimal energy waste, aligning well with the predicted values.

As concerns the visual comfort, Table 1 reports the results of our experiments during the four parts of the day. As we can see, in the morning, with an outdoor illuminance of 10000 lux, from the initial position (x=170, y=30, z=50), the desk lamp has to be lifted at the position (x=170, y=25, z=140) and it has to provide a luminous flux of 300 lumens to meet the user's target preference. By adopting these actions, the uniformity achieved is 0.761 and the work plane illuminance is 500 lux.

At noon, the outdoor illuminance increases to 50000 lux, owing to the sun's high position in the sky. The uniformity value of 1.0 indicates perfect distribution of light, meaning that the light is evenly spread across the work plane. The work plane illuminance of 1000 lux is easily achieved with the ambient daylight alone, meaning that no additional flux is needed from the desk lamp. In fact, the lamp remains in the same position, as natural light is sufficient to meet the required lighting levels indoors.

In the afternoon, with the outside illuminance decreasing to 5000 lux, the desk lamp again plays a more important role in supplementing the natural light.

The optimal position of the lamp shifts to (x = 170, y = 20, z = 150) and the uniformity value drops slightly to 0.699, indicating a somewhat less even light

distribution compared to noon. To achieve the target work plane illuminance of 500 lux the lamp needs to provide 400 lumens. This increase in luminous flux compared to the morning is due to the lower outdoor light levels as the sun moves toward the horizon.



Fig. 4. Energy waste comparison.

Finally, by the evening, outdoor illuminance has significantly reduced to 100 lux. As the contribution of natural light continues to diminish, the lamp must now provide 500 lumens of luminous flux to compensate the decreased ambient light and ensure the space remains adequately illuminated. To achieve the most effective lighting and uniformity for the work surface, the lamp should be moved to a more central position relative to the desk (x=135, y=25, z=150).

Time of Day	Outdoor Illuminance (lux)	Desk Lamp Position (x,y,z)	Desk Lamp Luminous Flux (lm)	Workplane Illuminance (lux)	Uniformity
Morning	10000	(170, 25, 140)	300	500	0.761
Noon	50000	(170, 25, 140)	0	1000	1
Afternoon	5000	(170, 20, 150)	400	500	0.699
Evening	100	(135, 25, 150)	500	502	0.692

Table 1. Results of RL Adaptation.

4. Discussions and Conclusions

In this paper, we presented a self-adaptive and self-learning lighting control system for energy efficiency and personalized comfort requirements.

At this stage, we have successfully realized the system's core functionality, which revolves around implementing and training two key components, the LSTM network, and the RL model, and integrating **these models** for real-time decision-making.

We are currently developing the whole prototype of our system, by using Python libraries such as Keras and TensorFlow to handle machine learning models and the TinyTuya Python library to interface with smart lamp. For data acquisition, we are using Delta Ohm HD 2021T (measuring range 0.02e20 klx) photosensors for monitor illuminance levels, enabling the system to adjust lighting based on ambient light conditions. The system is designed to control smart desk lamps, allowing for real-time adjustments of lighting settings, such as illuminance and on/off status, and to provide suggestions to the user for the most suitable desk lamp position. Then, the system will be deployed on a server to handle the control logic and data processing.

Regarding the computational aspects, currently, the core functionality of the systems runs on a **Mac Studio** with standard specifications (Apple M2 Max chip, 64 GB of RAM, 1 TB SSD with macOS Sonoma V.14.6). Thus, the training of the LSTM model takes

only a few minutes when using historical data. It's important to note that this training time is required only during the initial setup. Once the model is trained, it can be used for real-time predictions and adjustments without the need for retraining. At runtime, the system uses the trained LSTM model to process incoming data, and the complete sensing-control-adapt loop is achieved in just a few seconds.

No additional specialized hardware is required beyond the light sensors for data collection and the smart desk lamps for control. The system is designed to operate efficiently on the standard hardware mentioned, ensuring that it can function in real time with minimal latency.

Our next goal is to expand and refine this prototype into a comprehensive system that will incorporate user feedback to optimize lighting control in various environments. Additionally, we plan to conduct a series of experiments to evaluate the system's performance and gather data for post-assessment. These experiments will help us assess the effectiveness of our approach and identify areas for further improvement.

Acknowledgments

This work is funded by the European Commission – Next Generation EU – PNRR M4 – C2 -investimento 1.1 – PRIN 2022 cod. 2022YWW9B8 "Study for a tool for design, COntrol, and COmmissioning of Lighting Control systems. CUP Master: B53D23006660006.

References

[1]. Energy Efficiency 2018: Analysis and Outlooks to 2040, *International Energy Agency*, 2018.

- [2]. M. Beccali, M. Bonomolo, G. Ciulla, V. L. Brano, Assessment of indoor illuminance and study on best photosensors' position for design and commissioning of Daylight Linked Control systems. A new method based on artificial neural networks, *Energy*, Vol. 154, 2018, pp. 466-476.
- [3]. I. Konstantzos, S. A. Sadeghi, M. Kim, et al., The effect of lighting environment on task performance in buildings – A review, *Energy and Buildings*, Vol. 226, 2020, 110394.
- [4]. P. Kar, A. Shareef, A. Kumar, et al., ReViCEE: A recommendation-based approach for personalized control, visual comfort & energy efficiency in buildings, *Building and Environment*, Vol. 152, 2019, pp. 135-144.
- [5]. G. Ma, X. Pan, Research on a visual comfort model based on individual preference in China through machine learning algorithm, *Sustainability*, Vol. 13, Issue 14, 2021, 7602.
- [6]. C. Tzouvaras, et al., A novel dynamic approach for determining real-time interior visual comfort exploiting machine learning techniques, *Applied Sciences*, Vol. 13, Issue 12, 2023, 6975.
- [7]. G. Potenza, C. Baglivo, M. Bonomolo, P. Ribino, Optimizing photosensor placement for energy-efficient lighting in sustainable building design based on multivariate long short-term memory models, in *Proceedings of the 1st AIxIA Workshop on Green-Aware Artificial Intelligence* (GreenAI@AIxIA'24), 2024/
- [8]. EN 12464-1, Light and lighting Lighting of Work Places, Indoor Work Places, *European Committee for Standardization*, Brussels, Belgium, 2011.
- [9]. A. I. Slater, P. R. Boyce, Illuminance uniformity on desks: Where is the limit?, *Lighting Research and Technology*, Vol. 22, Issue 4, 1990, pp. 165-174.
- [10], H. F. O. Müeller, Sustainability, energy and architecture, Chapter 9, in Daylighting. *Academic Press*, 2013, pp. 227-255,

(040)

Generation of a Rhythm Descriptor in Musical Phrases Using Signal Processing and Artificial Intelligence Techniques

H. A. Aguilera-Garcia¹, R. A. Lizarraga-Morales²

 ¹University of Guanajuato, DICIS, Department of Multidisciplinary Studies, Av. Universidad S/n, Colonía Yacatitas, Yuriria, Gto., México
 ²University of Guanajuato, DICIS, Department of Art and Enterprise, Carr. Salamanca-Valle de Santiago 3.5+1.8 km Comunidad de Palo Blanco, 36885 Salamanca Gto., México E-mail: ha.aguileragarcia@ugto.mx, ra.lizarragamorales@ugto.mx

Summary: Nowadays, one of the most challenging tasks for very-large digital music datasets is their automatic management in terms of genre, mood, style, rhythm, and others. Such management is usually performed through metadata or descriptors of the musical features. The rhythm is one of the most notable features in music. However, its representation is still a challenge. It is desirable to extract it automatically and express it in the form of descriptors. In this work, the automatic computation of a descriptor of the rhythm is proposed. An experiment is conducted with recordings of a drum set containing performances of different musical genres. An energy based detection function and classifiers are used to identify musical notes onsets. The onsets are the raw material for the computation of Pairwise Variability Indexes that represents the irregularity in the rhythm. These indexes allow to organize the phrases based on the variability of the rhythm.

Keywords: Music information retrieval, Pairwise variability index, Short-time Fourier transform, Spectral flux detection function, Artificial intelligence.

1. Introduction

The Music Information Retrieval (MIR) research field aims to automatically manage large collections of digital audio content [1]. In order to classify a given music audio, different descriptors must be considered e.g. timbre, melody, rhythm, pitch, harmony, key, structure or lyrics. Given the massively increasing volumes of digitized music, the development of an automatic extraction of descriptors is an emerging need.

The rhythm is defined as the succession of sounds and silence over time. In musical compositions, the percussion instruments usually have the role of building what is called the rhythmic base, which determines the pulse of the musical piece. It is different with the other types of instruments, which are usually used to build melody or harmony.

In the existing literature, Shete and Deshmukh [2] propose to recognize five rhythmic patterns from North Indian music called Talas in the percussive instrument Tabla. The rhythm and rhythm-related aspects have been represented in other forms, such as indexes. metrics, statistics. or probabilities. Condit-Schultz [3] reviews the use of the Normalised Pairwise Variability Index (nPvi) in music to compare rhythmic patterns. It is remembered that it was originally proposed to compare rhythms in music with rhythms in speech. Chakraborty et al. [4], use several metrics to compare rhythms across languages, including Pairwise Variability Indexes (Pvi). The nPvi and the raw version, the Raw Pairwise Variability Index (rPvi). Panda [5] lists features extracted by toolboxes. These features are: Beat Spectrum, Beat Location, Onsets, Event Density, Average Duration of Events, Tempo, Metrical Structure, Metrial Centroid and Strength, Note Duration statistics, Note Duration Distribution, Ratios of Note Duration Transitions, Rhythmic Fluctations, Tempo Change and Rhythmic Clarity. Senn *et al.* [6] give probability values in a study to measure the complexity of rhythmic drum patterns. The probability correlates with the metrics: number of onsets, Syncopation Index, Kolmogorov Complexity and Revised Syncopation Index.

One of the most used indexes for the representation of rhythm is the Pvi. The Pvi represents the variability of the rhythm. A high Pvi value expresses an irregular rhythm, while a low Pvi indicates a constant and regular rhythm. The main difficulty in calculating the Pvi is the identification of onsets. It is desirable that this task can be performed automatically.

The activity of identifying musical note onsets has been studied widely in the past years. Recently, work has been done to improve identifying onsets. Gowriprasad and Murty [7] use linear prediction and Hilbert envelopes for onset identification. *Chen et al.* [8] present a Convolutional Neural Network (CNN) that processes 204 features. Mournir *et al.* [9] show the use of a detection function based on a normalized sparsity measure of spectrum magnitude. A detection function with an Echo State Network is created by Steiner *et al.* [10]. Tomczak and Hockman [11] identifies onsets using CNN and Bidirectional Temporal Neuronal Network. Kong *et al.* [12] suggest a regression-based onset identification system.

In percussive instruments, it is well known that the main characteristic of the signals at the moment of an onset is a significant increase in the amount of energy, Bello [13]. The elements of the previous literature show the possibility of identifying onsets and

representing rhythmic variability. However, the automation of the generation of rhythm descriptors from audio signals is an open task.

In this paper, musical phrases of a drum set performance are used to represent the rhythmic base of contemporary musical genres in the form of the rhythmic descriptors nPvi and rPvi. The Short-Time Fourier Transform (STFT) is used to extract spectral information from the audio signal and to build an energy based detection function. The values from the detection function are then used to train a classifier to identify onsets. The distances between onsets are the raw material for calculating the variability of the rhythm in the Pvi. This document is organized as follows. Section 1 contains the introduction. Section 2, the procedure for the calculation the indexes. Section 3, the results and analysis. Section 4, the corresponding conclusion.

2. Methodology

The computation of the Pvi values require eight steps, see Fig. 1. The experimental setup is inspired by Stasiak [14]. In the first step (Fig. 1a), recordings containing a musical phrase from 6 different musical genres, were taken from the Enst-Drums dataset [15]. The recordings are in WAV format, with a sampling frequency fs = 44.1 kHz. In the second step (Fig. 1b), the STFT was applied to time domain signals, is defined in Eq. (1).

$$X(n,k) = \sum_{m=-\frac{N}{2}}^{\frac{N}{2}-1} x(h+m)w(m)e^{-\frac{2j\pi mk}{N}}, \quad (1)$$

where x(m) is a point in the input signal, w(m) is a point in window, N is the window size = 2048, h is the hop = 441 y k is the frequency. In the third step (Fig. 1c), the spectral flux detection function (SF) is calculated as is shown in Eq. (2).

$$SF(n) = \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} H(|X(n,k)| - |X(n-1,k)|), \quad (2)$$

where H(x) = (x + |x|)/2 is half-wave rectifier function, |H(n,k)| is the module of k-bin in the current frame and |H(n-1,k)| is the module of the previous frame.



Fig. 1. The eights steps for the generation of the rhythm descriptor in musical phrases. a) Time domain signal. b) Short-time Fourier Transformation (STFT). c) Spectral flux detection function. d) Inputs preparation for classifier. e) Onsets identification. f) Time distances of onsets. g) Duration of note. h) Computation of rhythm descriptor Pvi.

In the fourth step (Fig. 1d), vectors of the shape Ve = [SF(n-2), SF(n-1), SF(n), SF(n+1),

SF(n+2)] were created with data points from the detection function, including the current, previous and following frames. The different vectors that we extract have cardinalities of $\{5, 7, 9\}$. An additional value \overline{SF} , the sum of 10 previous and following frames, has been added to others. These new vectors have cardinalities of $\{6, 8, 10\}$. All the resulting vectors were used as inputs to a classifier to determine the presence or absence of an onset in SF(n), the current frame. A total of 9,039 samples without onsets and 200 with onsets were available. Considering the problem of unbalanced classes, the class containing samples with onsets was oversampled with 2000 additional vectors. Such vectors were formed with 1000 copies of the existing ones, and 10 % noise was added to the other 1000.

In the fifth step (Fig. 1e), 5 types of classifiers were trained and validated in Weka [16] to identify onsets. The classifiers that we explore are the well-known ones: Bayesian Network (BN), Hoeffding Tree (HT),

Multilayer perceptron (MP), Decision Table (DT), 1-Nearest Neighbor (1-NN). We use 10-fold cross-validation for testing. In the sixth step (Fig. 1f), once the onsets are identified, the temporal distances between them are calculated. In the seventh step (Fig. 1g), these distances were normalized by taking the first distance as the standard. The first distance is assigned a value of 1. In the following ones values, the value 0.5 indicates that the distance is half of the first distance. If the value is 3, means 3 times the distance of the first one. The previous representation can be interpreted as the duration of the notes.

In the final step (Fig. 1h), the computation of the indexes PVI. The nPvi [3] and rPvi [4] are defined in Eqs. (3) and (4). The nPvi are integers in the range of 0 and 100. While the rPVI are floating-point numbers equal to 0 or greater.

$$nPvi = \left(\frac{100}{m-1}\right) \sum_{k=1}^{m-1} \left| \frac{d_k - d_{k+1}}{(d_k + d_{k+1})/2} \right|,\tag{3}$$

where *m* is the number of distances, d_k is the current distance between two onsets and d_{k-1} is the next.

$$rPvi = \sum_{k=1}^{m-1} \frac{|d_k - d_{k-1}|}{m-1}$$
(4)

3. Results and Analysis

The files of the musical phrases used for the experiment, their duration in seconds (D), and the number of onsets (O) are listed in Table 1. Notice that the difference in the number of onsets, reveals the differences in the rhythmic complexity among musical phrases.

Table 1. The files of the musical phrases, their duration in seconds (D), and the number of onsets(O).

File	D	0
036_phrase_disco_simple_slow_sticks	16	41
042_phrase_rock_simple_slow_rods	20	33
048_phrase_afro_simple_slow_mallets	10	15
060_phrase_salsa_simple_slow_sticks	20	45
066_phrase_shuffle-blues_simple_slow_brushes	14	35
078_phrase_reggae_simple_slow_sticks	13	31

The performance of the Bayesian network classifier is shown in Table 2. The metrics Accuracy (Acc), Precision (P), Recall (Rc), F-score (F1) are used to present the performance. |Ve| is the cardinality of the input vector. The best result is achieved with the 9-attribute vectors, with an accuracy of 94.4 %.

Table 2. Performance of the BN classifier.

Ve	Acc	Р	Rc	F1
5	0.850	0.901	0.850	0.861
6	0.840	0.890	0.840	0.852
7	0.938	0.943	0.938	0.940
8	0.936	0.943	0.936	0.938
9	0.944	0.948	0.944	0.945
10	0.940	0.946	0.941	0.942

The performance of the multilayer perceptron classifier is shown in Table 3. The best result is achieved with the 6-attribute vectors, with an accuracy of 97.6 %.

Table 3. Performance of the MP classifier.

$ V_e $	Acc	Р	Rc	F1
5	0.974	0.991	0.978	0.984
6	0.976	0.992	0.980	0.986
7	0.976	0.990	0.980	0.985
8	0.974	0.989	0.979	0.984
9	0.975	0.976	0.975	0.976
10	0.975	0.976	0.975	0.976

The performance of the decision table classifier is shown in Table 4. The best result is achieved with the 8-attribute vectors with an accuracy of 98.1 %.

The performance of the Hoeffding tree classifier is shown in Table 5. The best result is achieved with the 7-attribute vectors with an accuracy of 97.6 %.

Table 4. Performance of the DT classifier.

$ V_e $	Acc	Р	Rc	F1
5	0.966	0.966	0.966	0.966
6	0.975	0.976	0.976	0.976
7	0.979	0.980	0.980	0.980
8	0.981	0.982	0.982	0.982
9	0.979	0.980	0.980	0.980
10	0.978	0.978	0.978	0.978

Table 5. Performance of the HT classifier.

$ V_e $	Acc	Р	Rc	F1
5	0.963	0.965	0.964	0.964
6	0.973	0.974	0.973	0.974
7	0.976	0.978	0.976	0.977
8	0.973	0.974	0.974	0.974
9	0.974	0.976	0.975	0.975
10	0.975	0.976	0.975	0.975

The performance of the 1-nearest neighbor classifier is shown in Table 6. The best result is achieved with the 10-attribute vectors with an accuracy of 99.2 %.

Table 6. Performance of classifier 1-NN classifier.

$ V_e $	Acc	Р	Rc	F1
5	0.991	0.991	0.991	0.991
6	0.991	0.992	0.992	0.992
7	0.991	0.992	0.992	0.992
8	0.992	0.993	0.992	0.992
9	0.992	0.993	0.992	0.992
10	0.992	0.993	0.993	0.993

The best perfomances of each classifier are shown in Table 7. The classifier with the best result is 1-NN with an accuracy of 99.2 %.

Table 7. The best performances of each type of classifiers.

Classifier	$ V_e $	Acc	Р	Rc	F1
BN	9	0.944	0.948	0.944	0.945
HT	7	0.976	0.978	0.976	0.977
MP	6	0.976	0.992	0.980	0.986
DT	8	0.981	0.982	0.982	0.982
1-NN	10	0.992	0.993	0.993	0.993

Once the onsets in the musical phrases have been identified, the computation of the Pvi indexes is continued. The temporal distances between onsets for the afro genre phase are shown in Fig. 2 as example.

The previous temporal distances help to compute the duration of the notes. The note duration for the afro style musical phrase is shown in Fig. 3.

The values of the nPvi and rPvi of the musical phrases are ordered from the lowest to the highest variability are shown in Table 8. The rock genre phrase has the lowest variability and the afro genre phase has the highest variability.



Fig. 2. Time distance between the onsets in milliseconds for the musical phase of the file 048_phrase_afro_simple_slow mallets.wav.



Fig. 3. The duration of the notes for the musical phrase in the afro style in the file 048_phrase_afro_simple_allow_mallets.wav.

Table 8. The values of the nPvi and rPviof the musical phrases.

File	nPvi	rPvi
042_phrase_rock_simple_slow_rods	9.63	0.082
060_phrase_salsa_simple_slow_sticks	34.98	0.272
036_phrase_disco_simple_slow_sticks	35.22	0.496
066_phrase_shuffle- blues_simple_slow_brushes	50.64	0.404
078_phrase_reggae_simple_slow_sticks	57.98	0.410
048_phrase_afro_simple_slow_mallets	85.76	1.497

4. Conclusion

In this paper, the automatic generation of a rhythm descriptor (Pvi) is presented for digital music. Signal processing and artificial intelligence techniques are applied to recordings of musical phrases. High values of accuracy, precision, recall and F-score metrics are obtained for musical note detection of percussive instruments. These high values confirm that the energy based detection function highlights the moment when there is an onset. This corresponds to the increase in energy, which is a characteristic of percussive instruments. The nPvi and rPvi values represent the irregularity of the rhythm. The descriptors are means to perform the tasks of the MIR systems.

Acknowledgements

The support provided by Secretaria de Ciencia, Humanidades, Tecnología e Innovación (Secihti) from Mexico.

References

- M. A. Casey, R. Veltkamp, M. Goto, M. Leman, C. Rhodes, M. Stlaney, Content-based music information retrieval: Current directions and future challenges, *Proceedings of the IEEE*, Vol. 96, Issue 4, 2008, pp. 668-696.
- [2]. S. Shete, S. Deshmukh, North Indian classical music table tala (rhythm) prediction system using machine learning, in Advances in Speech and Music Technology, *Springer Nature*, 2021, pp. 187-197.
- [3]. N. Condit-Schultz, Deconstructing the NPVI: a methodological critique of the normalized pairwise variability index as applied to music, *Music Perception: An Interdisciplinary Journal*, Vol. 36, Issue 3, 2019, pp. 300-313.
- [4]. J. Chakraborty, L. Dihingia, P. Sarmah, R. Sinha, Effect of sociolinguistic variations on rate and rhythm of Hindi L2 speech, in *Proceedings of the Speech Prosody Conference*, Leiden, Netherlands, 2-5 July 2024, pp. 16-20.
- [5]. R. Panda, R. Malheiro, R. P. Paiva, Audio features for music emotion recognition: A survey, *IEEE Transactions on Affective Computing*, Vol. 14, 2023, pp. 68-88.
- [6]. O. Senn, F. Hoesl, R. Jerjen, T. A. Bechtold, L. Kilchenmann, D. Rose, E. Alessandri, A stimulus set of 40 popular music drum patterns with perceived complexity measures, *Music & Science*, Vol. 6, 2023.
- [7]. R. Gowriprasad, K. S. R. Murty, Onset detection of table strokes using LP analysis, in *Proceedings of the International Conference on Signal Processing and Communications (SPCOM'20)*, Bangalore, India, 19-24 July 2020. pp. 1-5.
- [8]. P.-H Chen, J.-J. Ding, J.-Y. Huang, T.-Y. Tseng, Accurate onset detection algorithm using feature-layerbased deep learning architecture, in *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS'20)*, Seville, Spain, 12-14 October 2020, pp. 1-5.
- [9]. M. Mounir, P. Karsmakers, T. V. Waterschoot, CNN-based note onset detection using synthetic data augmentation, in *Proceedings of the 28th European Signal Processing Conference (EUSIPCO'21)*, Amsterdam, Netherlands, 18-21 January 2021, pp. 171-175.
- [10]. P. Steiner, A. Jalalvand, S. Stone, P. Birkholz, Feature engineering and stacked echo state networks for musical ONET detection, in *Proceedings of the 25th International Conference on Pattern Recognition* (*ICPR'21*), Milan, Italy, 10-15 January 2021, pp. 9537-9544.
- [11]. M. Tomczak, J. Hockman, Onset detection for string instruments using bidirectional temporal and convolutional recurrent network, in *Proceedings of the* 18th International Audio Mostly Conference, Edinburgh, United Kingdom, 20 August – 1 September 2023, pp. 136-142.
- [12]. Q. Kong, B. Li, X. Song, Y. Wan, Y. Wang, High-resolution piano transcription with pedals by regressing onset and offset times, *IEEE/ACM Transactions on Audio, Speech and Language Processing*, Vol. 29, 2021, pp. 3707-3717.
- [13]. J. P. Bello, L. Daudet, S. Abdallah, C. Duxbury, M Davies, M. B. Sandler, A tutorial on onset detection in music signals, *IEEE Transactions on Speech and Audio Processing*, Vol. 3, Issue 5, 2005, pp. 1035-1047.

- [14]. B. Stasiak, J. Mońko, A. Niewiadomski, Note onset detection in musical signals via neural-network-based multi-ODF fusion, *International Journal of Applied Mathematics and Computer Science*, Vol. 26, Issue 1, 2026, pp. 203-213.
- [15]. O. Gillet, G. Richard, Enst-drums: an extensive audio-visual database for drum signals processing, in *Proceedings of the 7th International Society for Music*

Information Retrieval Conference (ISMIR'06), Victoria, Canada, 8-12 October 2006, pp. 156-159.

[16]. A. M. Chacón-Maldonado, G. Asencio-Cortés, F. Martínez-Álvarez, A. Troncoso, FS-studio: An extensive and efficient feature selection experimentation tool for WEKA explorer, *SoftwareX*, Vol. 23, 2023, 10140. (041)

Combined Feature Selection and Hyperparameter Optimization for Small Datasets

<u>N. L. Kämpf</u>^{1,2}

Berliner Hochschule für Technik, Department of Mathematics, Lütticher Straße 38., 13353 Berlin, Germany Freie Universität Berlin, Department of Information Systems, Garystraße 21, 14195 Berlin, Germany Tel.: +49 17652975227 E-mail: NickiLena.Kaempf@bht-berlin.de, nicki.kaempf@fu-berlin.de

Summary: This paper introduces a novel approach combining feature selection and hyperparameter optimization using Sequential Model-Based Optimization. Addressing the gap in small dataset applications, the method optimizes both features and hyperparameters simultaneously, mitigating issues like overfitting and the curse of dimensionality. The combined optimization is made feasible since features are treated as hyperparameters with two values: 0 for not selected and 1 for selected. In order to model both categorical and numerical hyperparameters, the Tree-structured Parzen Estimator is applied. The proposed method is tested on five datasets, demonstrating superior performance compared to traditional hyperparameter optimization or feature and model selection approaches. The results show better performance with less computation time.

Keywords: Feature selection, Optimization, Artificial intelligence, Machine learning, Small datasets.

1. Introduction

Small datasets are ubiquitous across various scientific disciplines, including bioinformatics, medicine and finance [1]. Despite their prevalence in practical applications, small datasets remain significantly underrepresented in the rapidly advancing research fields of Machine Learning (ML) and Artificial Intelligence (AI). The primary focus of ML and AI research has been on large-scale datasets, which, while beneficial for advancing algorithmic performance, do not reflect the constraints encountered in real-world scenarios. This discrepancy has resulted in a gap between theoretical advancements in ML and their practical deployment in industry settings.

Another critical issue is the limited focus on feature selection in research. Feature selection is essential for reducing model complexity, improving interpretability, and mitigating the risks associated with overfitting in scarce data settings [2]. However, most AI research emphasizes model architecture and optimization while neglecting the role of feature selection, further contributing to the difficulties in applying these models in practice.

To narrow the gap between scientific research and the application of ML and AI models in practice, this paper proposes a novel approach that combines feature selection with hyperparameter tuning using Sequential Model Based Optimization (SMBO). This integrated framework allows for an end-to-end optimization of ML and AI models, enabling them to perform effectively for small datasets. By leveraging SMBO optimize hyperparameters and features and simultaneously, the proposed method addresses key challenges such as overfitting and the curse of dimensionality. This work aims to provide a practical solution that enhances the generalizability and

applicability of ML and AI models, ultimately fostering the adoption of these models in domains where data is scarce.

2. Related Works

While hyperparameter tuning is well-researched, ranging from simple grid searches to elaborated SMBO approaches, research in the field of feature selection remains comparatively limited. Common feature selection methods include filter approaches, that evaluate the relevance of features based on statistical properties, such as mutual information or correlation, without considering the machine learning algorithm. and wrapper approaches, evaluate feature subsets by training a model and measuring its performance, with techniques like Recursive Feature Elimination or forward/backward selection [3]. Since the combination of multiple filter and wrapper approaches with numerous ML or AI models whose hyperparameter needs to be tuned results in a combinatorically demanding search space, AutoML algorithms like AUTO-Weka [4] or Auto-sklearn [5] leverage sophisticated optimization algorithms like SMBO to find the best combination of feature selection, model and hyperparameter set. However, both of these AutoML algorithm rely primarily on filter or wrapper approaches for the feature selection. Filter approaches fail to capture feature interactions. Moreover, relying on a measure for the statistical property of a feature is error-prone in itself, sensitive to the sample size and can struggle with noisy relationships [6,7]. Wrapper approaches select features based on model performance but are computationally expensive and prone to overfitting [3].

3. Theoretical Background for Sequential Model Based Optimization

SMBO includes a group of algorithms to optimize the hyperparameters for a given model. Thereby, SMBO is not limited to a specific model or a group of models but can be used for any model that needs to minimize any objective function. The general idea behind SMBO algorithms is that they take a pre-defined domain of hyperparameters and iteratively Since testing multiple sets test them. of hyperparameters can be time-consuming for complex models, the hyperparameter sets are evaluated on a surrogate function which is faster to calculate. The next hyperparameter set to test is found by taking into account the historic runs and the performance of the different hyperparameter values. In a nutshell, SMBO algorithms use sophisticated guesses to find the best set over a domain of hyperparameters while also reducing the run-time. In the following, the concept is explained in more detail to present the differences between the SMBO algorithms and justify the use of the Tree-Parzan Estimator algorithm in this paper.

All of the SMBO algorithms have 5 common parts: (i) a domain of hyperparameters to search over, (ii) an objective function that uses the hyperparameters and needs to be optimized, (iii) a surrogate function for the (cost-expensive) objective model, (iv) a selection function to decide on the next set of hyperparameters to test and (v) a history of the tested hyperparameters their performance. The domain and of hyperparameters consists of a probability distribution of continuous hyperparameter values and/or a list of discrete hyperparameter values. The probability distribution for each continuous hyperparameter needs to be selected by prior knowledge, which could be a source of error. However, the importance of hyperparameter search spaces for different ML and AI models have been extensively studied and best practices have been established. For example, studies have empirically proven that the minimum samples per leaf and maximal number of features for determining the split were the most important hyperparameters for Random Forests [8, 9].

The objective function is the loss function for the applied model. This loss will be optimized during the SMBO. Since the goal of this paper is to find the best set of input variables and hyperparameters, the possible combinations to test are huge und the computations are time consuming. Therefore, the number of combinations to test will be reduced by only calculating the loss function with the most promising combinations.

To find the most promising combinations, a surrogate function is built. The surrogate function is a probabilistic representation of the loss function given the hyperparameters. It maps the hyperparameter values to probability of a loss. To create the surrogate function and the probabilities, the SMBO algorithm needs to be started by some initial runs with random hyperparameter values. These values random hyperparameter values serve as a basis for the surrogate function. All of the surrogate functions use the Bayes Rule to model a loss probability y given a set of hyperparameters x:

$$p(y|x) = \frac{p(x|y) * p(y)}{p(x)}$$

The initial runs with random hyperparameters are used to create a prior distribution for p(y|x). The SMBO algorithms differ in the concrete application of Bayes Rule. In this paper, the Tree Parzan Estimator (TPE) will be used instead of the more frequently used Gaussian Processes. The empirical literature has proven that the TPE outperforms other surrogate functions such as Gaussian Processes [10, 3]. The TPE is also computationally less expensive since it scales linearly with the observations whereas the Gaussian Process scales cubically. Another distinction to Gaussian Processes is, that the TPE is able to handle categorical and especially binary variables directly. This feature of the TPE is crucial for the feature selection process developed in this paper and thus motivates the use of the TPE as a surrogate function.

The TPE creates two probability distribution for each hyperparameter: one distribution with hyperparameter values that resulted in a low loss (l(x))and one distribution with hyperparameter values that resulted in a high loss (g(x)). The separatio between a low and a high loss is done by a threshold y*. The relative frequency of runs performing better than the threshold y* determines p(y). Instead of directly modelling p(y|x), the TPE uses p(y) and p(x|y)given by:

$$p(x|y) = \begin{cases} l(x) \text{ if } y < y^* \\ g(x) \text{ if } y \ge y^* \end{cases}$$
(1)

The distributions of l(x) and g(x) are modelled by Parzen Kernel Density Estimators on the previous runs. Since the Parzen Kernel Density Estimators are the linear combination of Gaussian Mixture models, they also allow for a combination of categorical and continuous parameters. The tree structure mentioned in the name of TPE results of the tree-like hierarchy of the hyperparameters. This means that the parameters can be tuned step-by-step to account for hierarchies in the domain space. One example for a hierarchal domain space is i.e. if the number of units in each hidden layer is tuned individually, then the number of units in the third hidden layer only needs to be tuned if there exists a third layer.

On the basis of the surrogate function, the selection function finds the next set of hyperparameters to test on the loss function. There are different selection functions, with the most common choice being the Expected Improvement (EI). Combining the EI with the TPE yields the following equation:

$$EI_{y^*}(x) = \frac{\gamma y^* l(x) - l(x) \int_{-\infty}^{y^*} p(y) dy}{\gamma l(x) + (1 - \gamma)g(x)} \alpha \left(\gamma + \frac{g(x)}{l(x)}\right) (1 - \gamma)^{-1}$$
(2)

Intuitively, the EI should find the best next hyperparameter set to test. This is achieved by choosing hyperparameters that are more likely under the distribution l(x) than under the distribution from g(x). This is exactly what is promoted by Equation (2). The EI is higher, the higher the ratio between $\left(\frac{g(x)}{l(x)}\right)^{-1}$ as can be seen in the last term of Equation (2). In each iteration, many hyperparameter candidates are drawn from l(x) and evaluated by the EI. The hyperparameter set that maximizes the EI is chosen to be tested in the ML model and the loss function.

For the derivation, see Bergstra et al. [12].

With the next set of hyperparameters that are tested on the loss function, the distribution of l(x) and g(x)are updated. This process is repeated until a convergence criterion is met, i.e. the maximum number of iterations or the loss function has not improved for a predefined amount of iterations.

4. Methodology for Combined Feature Selection and Hyperparameter Optimization

Motivated by the drawbacks of the existing literature, a new algorithm called "Sequential Model Based Optimization with Feature Selection and Hyperparameter Tuning" (SMOFH) is proposed in this paper. This approach is similar to wrapper approaches. The difference is that the subset and number of features to include is simultaneously optimized with the models hyperparameters using SMBO. A schematic representation of the algorithm is shown in Fig. 1.



Fig. 1. Combined Feature Selection and Hyperparameter Optimization Approach SMOFH.

A ML or AI model is applied for either a regression or classification problem. The hyperparameters of the ML model are tuned with the SMBO algorithm described above. Up to this point this is common sense in applied ML. The new idea is to treat the features as 0/1 encoded hyperparameters in the SMBO. This means if a 0 is assigned to a feature, this feature will not be used in the next iteration. If a 1 is assigned to a feature, this feature will be used in the next iteration. In each iteration, the algorithm not only finds the next best hyperparameter set to test on the loss function but also the next best set of features to test.

This approach is in theory particularly effective, as it evaluates each feature independently, simultaneously optimizes features and hyperparameters. The SMOFH algorithm could potentially mitigate the curse of dimensionality posed by small datasets for ML and AI models by reducing the overall number of features as well as the problem of overfitting by only selecting the relevant ones.

5. Experiments and Results

To empirically evaluate the proposed methodology, it has been tested on five publicly available and diverse datasets and one proprietary one. The five publicly available datasets are binary or multiclass classification problems, whereas the proprietary dataset is a regression problem. Due to space constraints and the proprietary nature of the regression dataset, this paper focuses solely on the reproducible experiments and results from the five classification datasets.¹

Table 1. Overview Datasets.

	Lung [11]	Diabetes [12]	Credit [13]	Heart [14]	Student [15]
Classification problem	multi- class	binary	binary	binary	binary
Sample Size	27	520	659	1025	4424
# categorical features	56	15	9	8	19
# numerical features	0	1	6	51	17

An overview of the five classification datasets is given in Table 1. If there have been missing values in the dataset, these samples have been excluded for the analysis. For each dataset a random test-val-test split was applied with the ratio 0.7:0.15:0.15. For each of the datasets, the SMOFH has been applied to tune a Random Forest (RF) as well as a Multilayer Perceptron (MLP) for two scenarios: (1) for the proposed combined feature selection and hyperparameter optimization approach SMOFH and (2) if only a hyperparameter optimization was performed using SMBO. The loss function used in the RF and MLP training is the log cross entropy loss. The hyperparameter search space for the RF is based on findings of van Rijn and Hutter [8] that the maximal number of features for determining the split were the most important hyperparameters for Random Forests,

¹ Nevertheless, the SMOFH algorithm applied to a Long-Short-Term-Memory Network also outperformed other statistical, econometrical and ML models that have been

tuned and feature selection methods such as Lasso or Ridge for a time-series regression problem.

whereas the number of trees and the max depth control the complexity of the model. For the MLP, the hyperparameter search space is based on the finding that the learning rate can be considered the single-most important hyperparameter whereas the number of layers and number of units control the relation complexity that the MLP can learn [16]. The hyperparameter search space can be seen in Table 2.

 Table 2. Hyperparameter Search Space.

Hyperparameter Search Space RF					Hyperparameter Search Space MLP		
# Troos	# Trace May Donth		Min Samples		# Units	# Layers	Learning Rate
# Trees	Max Deptii	Leaf	Split		1 16		
100, 150, 200, 250, 300, 350, 400, 450	2-20	1, 2, 3, 4, 5	2, 3, 4, 5, 6		1, 16, 32, 48, 64, 80, 96, 100	1, 4, 7, 10	Log-uniform Sampling von $2 * 10^{-7}$ bis 10^{-2}

Additionally, the performance of the SMOFH algorithm has been compared to the feature and model selection approach for both, ensembles and single models, integrated in the Auto-sklearn, which has been the winning submission to the second ChaLearn AutoML challenge [4]. The Auto-sklearn 2.0 algorithm covers different filter approaches for the feature selection, for example filter approaches based on variance thresholding and selecting the k-best features based on the Mutual Information Criterion, Feature Importance or Chi-Square tests as well as feature reduction techniques such as Principal Component Analysis. Moreover, the Auto-sklearn 2.0 algorithm covers a variety of state-of-the-art classification models, such as tree-based models, linear models with integrated feature selection approaches such as Ridge or Lasso, Bayesian Models, MLP as well as Support Vector Machines and k-nearest Neighbor Classifier. Auto-sklearn 2.0 was chosen over AUTO-Weka due to its wider adoption and better maintenance. improving reproducibility. The maximum compute time for the Auto-sklearn 2.0 algorithm is set to 10 minutes to yield a fair runtime comparison to the SMOFH algorithm.

All experiments were conducted on a Lenovo Yoga C740, featuring an Intel Core i7-10510U CPU with 4 cores and a clock speed of up to 2.3 GHz. The system is equipped with 16 GB of DDR4 RAM and a 512 GB SSD for storage. The operating system used was Windows 11 Pro (version 23H2), with the WSL distribution Ubuntu 22.04 LTS and the Linux 5.10.xx kernel. The python libraries hyperopt 0.2.7, scikit-learn 0.24.2 along with Auto-sklearn 2.00.15.0 were employed for the computational tasks¹.

To compare the performance, each model predictions of the test set are evaluated using the log cross entropy loss (log loss), the accuracy as well as the Receiver Operating Characteristic Area Under Curve (ROCAUC) score. Thereby, the distinct advantages of each metric can be incorporated into a comprehensive evaluation of the model performance, such as favorable characteristics for balanced datasets (accuracy), unbalanced datasets (log loss) and independence of a specific classification threshold (ROCAUC).

The results for the RF can be found in Table 3. In Table 3 only the best RF for the proposed SMOFH and the best RF under the traditional SMBO for hyperparameter tuning are shown. The best RF are selected based on the lowest log loss on the validation set. Across the four smallest datasets (Lung, Diabetes, Credit, Heart), the proposed SMOFH algorithm performs as well or outperforms the traditional SMBO approach where only the hyperparameters are tuned as well as the ensemble and single model tuned under the Auto-sklearn 2.0 algorithm. This is not only true for the best optimization iteration, but also in the vast majority of cases comparing the 5, 10, 20 or 30 best iterations. This underlines the fact that the proposed SMOFH algorithm does not lead to one lucky configuration that outperforms other models but that a combined feature selection and hyperparameter optimization leads to overall better results for small datasets. For the largest dataset (Student), the proposed SMOFH algorithm did outperform the traditional SMBO, but not the Auto-sklearn algorithm.

Moreover, the computation time is shorter for the proposed SMOFH algorithm than the other algorithms. This is due to the fact that the feature selection optimization results in better performances in early optimization iterations and no performance increase can be made, thus triggering the early stopping criterion in fewer iterations. These finding underline the importance of integrating a designated, individual feature selection into the optimization process of ML and AI models.

Due to the page limit the results for the MLP are not explicitly stated. To summarize the findings regarding the MLP, the SMOFH algorithm consistently outperforms the traditional SMBO approach where only the hyperparameters are tuned. However, the performance of the MLP tuned with the

¹ For reproducability reasons, the random seed for the train-val-test split in scikit-learn was set to 0. For hyperopt and auto-sklearn, all random seeds were set to 1.

SMOFH algorithm is inferior to the ensemble and single models trained. This is in line with findings in the literature that MLP perform worse than ML models

for small datasets [16]. This has also been shown for the lung dataset [12] and the credit dataset [14] used in this analysis.

Table 3. Algorithm Performance	Comparison,	bold numbers	highlight the	best performance	within a dataset.
--------------------------------	-------------	--------------	---------------	------------------	-------------------

	Algorithm	Log loss	ROCAUC	Accuracy	Run time
Lung	proposed SMOFH	0.7439	0.8333	0.8000	3:52
	Traditional SMBO	0.9321	0.8333	0.6000	12:48
	Auto-sklearn ensemble	0.7956	0.8889	0.6000	10:00
	Auto-sklearn single	0.7305	1.0000	0.6000	10:00
	proposed SMOFH	9.9920 <i>e</i> ⁻¹⁶	1.0000	1.0000	2:09
Diabotos	Traditional SMBO	1.3284	0.9586	0.9615	4:35
Diabetes	Auto-sklearn ensemble	17.2697	0.4333	0.5000	10:00
	Auto-sklearn single	14.6129	0.5000	0.5769	10:00
	proposed SMOFH	4.1866	0.8727	0.8788	4:03
	Traditional SMBO	4.1865	0.8705	0.8788	7:22
Creun	Auto-sklearn ensemble	4.5354	4.5354	0.8687	10:00
	Auto-sklearn single	4.8843	4.8843	0.8586	10:00
	proposed SMOFH	6.0068	0.8207	0.8261	1:24
Hoort	Traditional SMBO	7.5085	0.7758	0.7932	1:28
mart	Auto-sklearn ensemble	6.7577	0.7865	0.8043	10:00
	Auto-sklearn single	7.5085	0.7680	0.7826	10:00
Student	proposed SMOFH	5.2537	0.7890	0.8479	14:26
	Traditional SMBO	5.2537	0.7806	0.8479	33:10
	Auto-sklearn ensemble	4.3173	0.8308	0.8795	10:00
	Auto-sklearn single	4.7855	0.8224	0.8614	10:00

6. Discussion

In the following, potentially imitating factors for this study as well as the generalization ability of the experiments will be discussed.

The choice of the hyperparameter search space in SMBO could significantly influence the performance of the optimization process. In this study, the hyperparameter search space was carefully selected based on established scientific findings, ensuring that even with a limited number of hyperparameters, the proposed method was able to outperform benchmark models. This demonstrates that an informed selection of the search space can lead to an effective and efficient optimization process. Beyond the search space itself, the performance of SMBO is also affected by the settings of its own hyperparameters. In this study, we set the maximum number of trials to 1000 and the early stopping criterion to 150 trials without improvement. Interestingly, these settings did not seem to have a significant impact on the final performance, as the proposed SMOFH algorithm reached a solution that outperformed other models well before the 1000 trials limit. Additionally, an ablation study was conducted by increasing the early stopping criterion to 300 trials without improvement. The results showed no difference compared to the setting with 150 trials, suggesting that an even lower early stopping threshold could yield similarly strong results. This indicates that

the proposed method is robust and can achieve optimal solutions without requiring excessively long optimization processes.

A key advantage of the SMOFH algorithm is its capability to provide implicit feature importance. Unlike traditional feature importance methods, which require separate post-hoc analyses, implicit feature importance is derived directly from the optimization process. This approach offers several benefits. Since feature importance is inferred during the optimization, no additional computational steps are required, reducing overall runtime. The importance scores are inherently aligned with the optimization process, ensuring that they reflect the actual contribution of each feature to the model's performance. By analyzing the frequency and influence of specific features during the optimization, researchers can gain insights into which features are most relevant for the task, potentially guiding further feature engineering efforts. Implicit feature importance is less prone to biases introduced by separate importance estimation methods, as it is integrated into the optimization framework itself.

While there is no unequivocally definition for small datasets in the ML or AI research, a literature review from Rather et al. in 2024 found that for the majority of studies, datasets with less than 3000 samples per class are defined as small datasets [17]. With this rule of thumb, the four smallest datasets (Lung, Diabetes, Credit, Heart) in this study can be classified as small. For these datasets, the proposed SMOFH algorithm was able to outperform all other benchmark models, including the ones trained via the Auto-sklearn 2.0 algorithm. The largest dataset examined in this study, the student dataset, has 3003 instances in the class "Graduated" and 1421 instances in the class "Dropout". While Rather et al. found that a minority of papers also label datasets with more than 3000 samples per class as small [17], the SMOFH algorithm was not able to outperform the models trained with the Auto-sklearn 2.0 algorithm. While the RF trained with SMOFH did not outperform the models from the Auto-sklearn 2.0 algorithm, there are several potential explanations for this result. One possible reason is that the RF suffers from overfitting. For all metrics, the RF shows significantly better results on the validation set than on the test set. When comparing the metrics achieved by the RF optimized with the proposed SMOFH algorithm on the validation data, it actually outperformed the Auto-sklearn 2.0 algorithm on the test data, indicating a potential discrepancy between training and testing phases. This could point to a peculiarity in the train-validation-test split or a specific characteristic of the student dataset that influenced the model's performance.

However, even with datasets with slightly more than 3000 samples per class, the presented SMOFH algorithm was able to outperform traditional SMBO approaches for hyperparameter tuning. This suggests that the combined feature selection and hyperparameter optimization process used by SMOFH could have significant benefits for larger datasets as well, making it potentially valuable for optimizing deep learning models on larger, more complex datasets.

In light of these observations, further research and empirical experiments are needed to explore the importance and effectiveness of combined feature selection and hyperparameter optimization on larger datasets. Specifically, it would be valuable to investigate whether the benefits demonstrated on smaller datasets hold when applied to more complex problems or whether the advantages of SMOFH are more pronounced for certain types of datasets or machine learning models. Such research would be crucial in understanding the broader applicability of SMOFH and its potential impact on optimizing machine learning models for a wide range of applications.

7. Conclusion

In conclusion, the proposed SMOFH algorithm demonstrates significant improvements in model performance, particularly in small dataset scenarios. By simultaneously optimizing the hyperparameters and the individual feature selection, the method outperforms traditional hyperparameter optimization techniques and AutoML algorithms while also reducing computational costs. Empirical results across various classification datasets highlight the SMOFH algorithm's ability to deliver robust performance in resource-constrained environments. These findings highlight the practical applicability of SMOFH in domains such as bioinformatics, medicine, and finance, where small datasets are common.

Future research should focus on expanding the SMOFH approach to larger, more complex datasets, and explore its potential for optimizing deep learning models, offering broader insights into its applicability. Further studies on the integration of feature selection with advanced optimization techniques could lead to enhanced AutoML systems with even better performance and efficiency. Therefore, more empirical evidence is needed to give empirical insights for the, choice of ML and AI models, the predefined hyperparameter space, the maximum number of optimization iterations and the early stopping criterion.

References

- J. Fan, R. Li, Statistical challenges with high dimensionality: Feature selection in knowledge discovery, in *Proceedings of the International Congress of Mathematicians*, Vol. 3, 2006, pp. 595-622.
- [2]. I. Guyon, A. Elisseeff, An introduction to variable and feature selection, *Journal of Machine Learning Research*, Vol. 3, 2023, pp. 1157-1182.
- [3]. L. Talavera, An evaluation of filter and wrapper methods for feature selection in categorical clustering, in *Proceedings of the International Symposium on Intelligent Data Analysis*, 2005, pp. 440-451.

- [4]. C. Thornton, C. F. Hutter, et al., AUTO-Weka: Combined selection and hyperparameter optimization of classification algorithms, in *Proceedings of the 19th* ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2013, pp. 847-855.
- [5]. M. Feurer, K. Eggensperger, et al., Auto-sklearn 2.0: Hands-free AutoML via meta-learning, *Journal of Machine Learning Research*, Vol. 23, 2022, pp. 11936-11996.
- [6]. K. Hopf, S. Reifenrath, Filter methods for feature selection in supervised machine learning applications – review and benchmark, *arXiv preprint*, 2021, arXiv:2111.12140.
- [7]. E. O. Abiodun, A. Alabdulatif, et al., A systematic review of emerging feature selection optimization methods for optimal text classification: the present state and prospective opportunities, *Neural Computing and Applications*, Vol. 33, 2021, pp. 15091-15118.
- [8]. J. N. Van Rijn, F. Hutter, Hyperparameter importance across datasets, in *Proceedings of the 24th ACM* SIGKDD International Conference on Knowledge Discovery & Data Mining, 2018, pp. 2367-2376.
- [9]. P. Probst, N. Marvin, et al., Hyperparameters and tuning strategies for random forest, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, Vol. 9, Issue 3, 2019, e1301.
- [10]. J. Bergstra, J., R. Bardenet, et al., Algorithms for hyperparameter optimization, in Advances in Neural

Information Processing Systems, Vol. 24, Curran Associates, Inc., 2011.

- [11]. Z. Hong, J. Yang, Lung Cancer [Dataset], UCI Machine Learning Repository, https://archive.ics.uci.edu/dataset/62/lung+cancer
- [12]. Early Stage Diabetes Risk Prediction [Dataset], UCI Machine Learning Repository, https://archive.ics.uci.edu/dataset/529/early+stage+dia betes+risk+prediction+dataset
- [13]. J. Quinlan, Credit Approval [Dataset], UCI Machine Learning Repository, https://archive.ics.uci.edu/ dataset/27/credit+approval
- [14]. D. Lapp, Heart Disease Dataset, Kaggle, https://www.kaggle.com/datasets/johnsmith88/heartdisease-dataset?resource = download
- [15]. V. Realinho, M. Vieira Martins, et al., Predict Students' Dropout and Academic Success [Dataset] UCI Machine Learning Repository, https://archive.ics.uci.edu/dataset/697/predict+student s+dropout+and+academic+success
- [16]. H. Xu, K. Kinfu, et al., When are deep networks really better than decision forests at small sample sizes, and how?, *arXiv preprint*, 2021, arXiv:2108.13637.
- [17]. I. H. Rather, S. Kumar, et al., Breaking the data barrier: a review of deep learning techniques for democratizing AI with small datasets, *Artificial Intelligence Review*, Vol. 57, 2024, 226.

(042)

Res-Scrum: A Proactive and Resilient Agile Framework for Managing Uncertainty in Software Development

Aziz Fellah

Northwest Missouri State University, School of Computer Science and Information Systems, Maryville, MO 64468 USA E-mail: afellah@nwmissouri.edu

Summary: This paper introduces Res-Scrum (Residual Scrum), a proactive and resilient Agile framework that strengthens Scrum by integrating residuality principles. Res-Scrum introduces key elements – residual sprint, residual backlog, residual controller, and residual retrospective – which expand Scrum's capabilities by adding adaptivity, proactivity, and resilienceoriented components. These elements empower development teams to anticipate and manage unexpected challenges and uncertainties throughout the software development lifecycle. By embedding residuality into sprints, backlogs, and retrospectives, Res-Scrum reinforces Scrum's capacity to withstand uncertainty, stressors, and challenges. This layered approach establishes a forward-thinking methodology that enables teams to build systems that are both adaptive and resilient to emerging complexities. A case study showcases its effectiveness in developing a resilient smart home automation app with proactive risk management.

Keywords: Residual scrum, Agile framework, Proactive resilience, Residuality theory, Adaptive software development, Stressor, Uncertainty.

1. Introduction and Related Work

In the evolving field of software engineering, agility and adaptability are crucial for managing complexity and uncertainty. Scrum, a widely adopted Agile framework, structures iterative development through roles, events, and artifacts, promoting close collaboration. However, despite its strengths, Scrum often struggles with unexpected disruptions and uncertainties, which can threaten project outcomes.

To address these challenges, we introduce Res-Scrum (Residual Scrum), the first framework to integrate Residuality Theory into Scrum, enhancing its resilience and adaptability. Res-Scrum focuses on presimulating potential stressors and uncertainties throughout the software development lifecycle, allowing teams to respond adaptively to unforeseen challenges. Unlike predictive approaches, Res-Scrum uncovers emergent stable behaviors (attractors) instead of relying on exhaustive scenario planning.

The framework introduces a dual-sprint system with regular and residual backlogs, ensuring both planned tasks and unforeseen complexities are effectively managed. Res-Scrum also incorporates residual retrospectives to address unresolved challenges and a residual controller role to oversee stressor mitigation. By embedding residuality principles into Scrum, Res-Scrum strengthens the framework's ability to navigate unpredictable environments while maintaining the core agile values of collaboration, flexibility, and continuous improvement.

1.1. Agile, Scrum and Uncertainty

Agile and Scrum [1, 16-19] are well-studied in software engineering, particularly for their impact on

team collaboration, efficiency, and adaptability within the Software Development Life Cycle (SDLC). While Scrum embraces uncertainty through iterative development, managing unpredictability remains challenging due to evolving requirements and dynamic stakeholder expectations This paper explores the core principles of Res-Scrum and demonstrates its potential to enhance the resilience and sustainability of software development practices in dynamic and. complex environments.

Research on Agile and Scrum covers areas such as team effectiveness, resilience, and large scale implementations. Five key factors influencing Scrum team performance, including responsiveness and stakeholder engagement have been studied in [2, 6]. A resilience-based agility model for navigating disruptions has been proposed in [5]. Other studies compare Agile with traditional project management and examine Scrum in safety-critical systems Architectural uncertainty, 8. 11. 16-18]. [7, particularly in design decisions, has received limited focus. Literature reviews [1, 3] emphasize adaptive design strategies to handle uncertainty in software architecture. While Scrum's iterative nature helps manage uncertainty, major disruptions still present challenges [3, 4, 7-9, 11].

1.2. Residuality Theory and Its Role in Scrum

Residuality Theory, introduced by O'Reilly and others [5, 9, 12, 13], addresses persistent uncertainties in complex systems. Unlike traditional approaches that eliminate uncertainty, Residuality Theory focuses on pre-simulating challenges and designing systems to stressors. This aligns with Agile and Scrum's core principles of adaptability and responsiveness. Despite its strengths, Scrum struggles with systemic uncertainties. While effective for sprint-level adjustments, it lacks mechanisms for proactively managing long-term disruptions. Integrating residuality principles into Scrum could enhance its resilience, improving adaptability in dynamic environments.

This paper is structured as follows: Section 2 highlights the core principles of Res-Scrum in software development and the importance of Residuality Theory in managing uncertainty. Section 3 introduces Res-Scrum, explaining how it integrates residuality by redefining roles, events, and artifacts to enhance resilience. Section 4 details the residual controller, a key role for managing complexity and unpredictability. Section 5 outlines Res-Scrum's core principles, reinforcing Scrum's adaptability. Section 6 provides guidelines for integrating regular and residual concepts into Agile practices. Section 7 provides some insights into implementing Res-scrum in Agile. Section 8 concludes with key insights on Res-Scrum's impact. Due to space constraints, the full details of the case study are available upon request from the author.

2. Core Principles of Res-Scrum

Res-Scrum builds on traditional Scrum values while incorporating residuality to enhance resilience and adaptability. It adds a layer for managing uncertainties, allowing teams to anticipate and prepare for disruptions rather than react to them. Residuality doesn't predict or control future changes but identifies stable system patterns using stressor simulations to uncover hidden relationships. It embraces unpredictability, focusing on the system's inherent behavior rather than trying to define a rigid architecture. Key principles include:

2.1. Proactive Anticipation

Res-Scrum encourages teams to foresee disruptions and prepare contingency plans, allowing them to address challenges early, ensuring smooth project progression and reducing the need for urgent adjustments.

2.2. Residual Resilience

Res-Scrum builds resilience by ensuring teams and systems can withstand and adapt to unforeseen disruptions, maintaining stability and quality through challenges. This principle ensures that systems remain functional and effective even in uncertain environments.

2.3. Residual Antifragility

Inspired by Taleb's antifragility [9, 14], Res-Scrum enables systems not only to survive disruptions but to thrive and improve because of them. Unlike resilient systems that endure, antifragile systems grow stronger through uncertainty. Every sprint in Res-Scrum is a chance to learn and evolve, with stressors serving as opportunities for growth.

3. Residuality in Scrum: Adapting Roles for Resilience and Agility

Res-Scrum refines scrum roles to improve resilience, manage disruptions, and adapt to evolving needs. It introduces the residual product owner, residual sprints, residual backlog, and residual controller while adapting events and artifacts for proactive risk management. The product owner anticipates disruptions, the scrum master fosters adaptability, and the development team builds resilient solutions. The residual controller oversees stressors and adjusts priorities. Together, these roles ensure robust, adaptable project outcomes.

3.1. Residual Sprints vs. Regular Sprints: A Dual-Layered Framework

In Res-Scrum, regular and residual sprints serve distinct yet complementary roles. Regular sprints on delivering product increments - new features and improvements - guided by the product backlog in a standard 1-4 week cycle. Residual sprints, typically 1-2 weeks long, address risks, technical debt, and unresolved challenges that impact long-term success. Driven by the residual backlog, they tackle architectural issues, manage complexities, and enhance adaptability without disrupting primary development. Fig. 1, and Fig. 2 illustrate the distinction: one showing a standard product backlog with regular sprints, the other highlighting a residual backlog with residual sprints. By balancing immediate progress with proactive risk management, Res-Scrum ensures both stability and long-term resilience.

Fig. 1 and Fig. 2 compare the product backlog and residual backlog, showing the management of regular and residual sprints with integrated residuality in the scrum master role.



Fig. 1. Scrum framework.

3.2. Residuality Backlogs vs. Product Backlogs

A residual backlog focuses on risks, unresolved complexities, and future uncertainties, the product
backlog, which prioritizes immediate tasks. By addressing long-term resilience, it helps teams proactively manage challenges. Fig. 1 and Fig. 2 illustrate the product backlog with regular sprints in Scrum and the residual backlog with residual sprints in Res-Scrum.

Residual Backlog		Residual Sprint 1	Residual Sprint 2	Residual Sprint
Risks	Technical Debts	X1	X1	X1
Un- certainties	Emergent Challenges			

Fig. 2. Res-Scrum framework.

3.3. Scrum Master and Residuality

Residuality integrates seamlessly into the Scrum Master role, eliminating the need for a separate position. This expanded role includes proactive risk management, guiding the team through uncertainties while maintaining resilience. By combining regular and residual responsibilities, the Scrum Master ensures productivity, adaptability, and risk awareness without added complexity (see Fig. 3).



of Scrum Master.

3.4. Residual vs. Regular Retrospectives

Regular retrospectives refine workflows by evaluating past sprints, while residual retrospectives assess how well the team anticipates and manages stressors. As shown in Fig. 4, this dual approach balances immediate improvements with long-term resilience in Res-Scrum. Similar to the duality of sprints and backlogs illustrated in Fig. 1, we divided scrum retrospectives into two categories, as shown in Fig. 4, regular retrospectives, which align with Scrum traditional practices, and residual retrospectives, a distinctive feature of Res-Scrum that focuses on addressing residual complexities and unresolved issues

3.5. Product Owner and Residuality

The product owner retains almost all traditional responsibilities while incorporating risk management,

disruption anticipation, and alignment with evolving client needs. This integrated approach balances immediate priorities with long-term resilience, ensuring adaptability without adding complexity.

3.6. Enhanced Events and Artifacts

Res-Scrum extends Scrum to address stressors, uncertainties, and risks while main- training adaptability. Events like daily stand-ups and sprint reviews now emphasize resilience, and artifacts such as backlogs prioritize risk mitigation. By embedding residuality, Res-Scrum helps teams anticipate challenges, sustain progress in uncertainty, and balance Agile practices with proactive risk management.



Fig. 4. Res-Scrum Residual Retrospective vs. Scrum Regular Retrospective.

4. Residual Controller Management

The residual controller is a key role in managing complexities and uncertainties within iterative development. Acting as a stabilizing force, it oversees residual components, implements stressor mitigation strategies during sprints, and monitors the residual backlog for potential risks. In collaboration with the scrum naster and development team, the residual controller prioritizes resilience tasks, ensuring the team remains adaptable and prepared for unforeseen challenges. The diagram below illustrates the residual controller's connections with other roles and project components. Fig. 5 shows the cycle of residual scrum interactions.



Fig. 5. Cycle of Residual Scrum Interaction.

5. Integrating Regular and Residual Agile Practices

The following set of sections provide clear guidelines for integrating regular and residual concepts within Agile practices, ensuring teams address both immediate delivery needs and long-term system resilience effectively.

5.1. Comparison of Regular and Residual Sprints

Dividing sprints into regular and residual sprints aligns with Residuality Theory, providing a structured approach to feature development while addressing accumulated complexities and risks. Below are suggestions for optimizing this framework.

Regular Sprints:

- Objective: Develop and deliver new features aligned with the product backlog;
- Priority: Focus on user stories that directly contribute to product goals;
- Ceremonies: Follow standard Scrum practices (e.g., planning, stand-ups, retrospectives).

Residual Sprints:

- Objective: Resolve complexities, reduce technical debt, and address risks;
- Pre-Simulation: Anticipate stressors and assess residual complexities;
- Priority: Tackle items from the "Residual Backlog" (deferred or flagged tasks);
- Residuals: Identify risks, unresolved issues, and technical debt;
- Risk and Lessons: Improve architecture, resolve technical debt, and integrate lessons into future sprints.

5.2. Structuring Backlogs for Sustainable Development

Separating backlogs into regular and residual types helps teams balance feature delivery with system resilience. This approach manages technical debt and risks without disrupting development while simplifying tracking and prioritization of immediate needs versus long-term stability.

Regular Backlog:

- Objective: Focuses on features, user stories, and tasks aligned with product goals, prioritizing customer-driven enhancements;
- Ownership: Managed by the Product Owner with input from the development team.

Residual Backlog:

- Objective: Ensures system resilience by tracking unresolved issues, technical debt, and non-feature tasks from regular sprints;
- Ownership: Overseen by the residual controller, working with the scrum master and development team to address risks and maintain long-term adaptability.

5.3. Regular vs. Residual Retrospectives

Regular retrospectives focus on overall team feedback, while residual retrospectives target specific areas like delivery processes or risk management. Adjusting metrics for each type helps track progress more effectively and drives continuous improvement.

Regular Retrospectives:

- Objective: Develop and deliver new features aligned with the product backlog;
- Focus on overall sprint feedback, team collaboration, and process improvements for feature development;
- Ownership: Led by the scrum team, including the scrum master and product owner.

Residual Retrospectives:

- Assess residual management, including risks, technical debt, and complexities, by analyzing pre-simulation results and mitigation strategies;
- Ownership: Managed by the residual controller and team members, ensuring effective risk tracking and long-term resilience.

6. Managing Dual Sprints: A Guide to Agile Coordination

This section provides strategies for synchronizing dual sprints in Agile, ensuring smooth workflows, collaboration, and reduced technical debt from short-term delivery trade-offs.

Residual Backing:

- Keep a separate residual backlog for technical debt, risks, and unresolved tasks;
- Prioritize based on impact, risk, and project goals;
- Review items during each sprint planning session.

Residual Sprint Scheduling:

- Alternate between regular and residual sprints (e.g., three regular sprints followed by one residual);
- Use residual sprints as needed when residuals reach specific thresholds.

Residual Sprint Planning:

- The scrum master (or residual master) leads planning;
- Analyze root causes, assign mitigation tasks, and set goals for testing or refinement.

Metrics for Residual Sprints:

- Residual burn-down: Track resolved residual issues;
- Risk Mitigation Score: Measure the reduction of risks;
- Technical Debt Reduction: Assess improvements in code quality;
- Retrospectives: Focus on the impact of residual work on project resilience.

7. Implementing Res-Scrum in Agile Workflows

Res-Scrum integrates residual sprints for stressors and regular sprints for progress, with retrospectives balancing immediate and long-term strategies for resilience and risk management.

Backlog Management:

- Cross-Link Backlogs: Connect regular and residual items (e.g., a feature backlog item may generate a residual task like refactoring);
- Prioritization: Rank residual items by impact and urgency;
- Regular Reviews: Align both backlogs with project goals during sprint planning.

Retrospective Enhancements:

- Focused Discussions: Regular retrospectives address delivery and efficiency; residual retrospectives analyze risks and mitigation;
- Analytical Tools: Use fishbone diagrams or 5 Whys for root cause analysis; encourage brainstorming for pre-simulation improvements. Integration across Sprints:
- Regular sprints: Drive feature development and process optimization;
- Residual sprints: Strengthen system resilience and reduce technical debt;
- Visibility and Communication: Share residual sprint outcomes with stakeholders;
- Continuous Adaptation: Periodically review and refine the dual-sprint framework for optimal balance.

8. Conclusion

This paper introduces Res-Scrum, an enhanced Scrum framework that integrates residuality to boost resilience and adaptability in software development. By redefining roles, events, and artifacts – such as the residual product owner, sprints, and controller – Res-Scrum helps teams proactively manage uncertainties. It extends scrum's iterative approach with residual components to anticipate disruptions and maintain flexibility.

Balancing regular and residual practices enables teams to meet immediate delivery needs while ensuring long-term resilience. A smart home automation case study illustrates how Res-Scrum fosters adaptability and effective management of planned and unforeseen requirements. The structured division of backlogs, retrospectives, and sprint reviews sharpens focus, systematically integrating risk and resilience into Agile workflows.

However, implementing Res-Scrum presents challenges. Overlapping priorities between backlogs can create confusion, but clear categorization guidelines ensure that tasks improving functionality belong in the regular backlog, while those reducing future risks or enhancing resilience go into the residual backlog.

Acknowledgements

This research was supported by funding from Northwest Missouri State University. The author would like to express his gratitude to the university for its support in facilitating this study.

References

- T. N. F. Pereira, M. S. de Oliveira, et al., Scrum: A systematic literature review, *International Journal of Advanced Computer Science and Applications*, Vol. 14, Issue 4, 2023, pp. 173-181.
- [2]. C. Verwijs, D. Russo, A theory of scrum team effectiveness. ACM Transactions on Software Engineering and Methodology, Vol. 32, Issue 3, 2023, pp. 1-51.
- [3]. C. Lupafya, A framework for managing uncertainty in software architecture, in *Proceedings of the 13th European Conference on Software Architecture* (*ECSA'19*), Paris, France, for Computing Machinery, 2019, pp. 71-74.
- [4]. C. Lupafya, D. Balasubramaniam, A framework for considering uncertainty, in *Proceedings of the IEEE* 46th Annual Computers, Software, and Applications Conference (COMPSAC'22), 2022, pp. 1519-1524.
- [5]. M. Lotfi, M. S. Sodhi, Resilient agility under the practice-based view, *Production Planning & Control*, Vol. 35, Issue 7, 2024, pp. 670-682.
- [6]. M. Barbareschi, S. Barone, Automatic test gene-ration to improve scrum for safety agile methodology, in *Proceedings of the 18th International Conference on Availability, Reliability and Security (ARES'23)*, 2023, pp. 1067-1088.
- [7]. D. Donmez, G. Grote, The practice of not knowing for sure: How agile teams manage uncertainties, in *Proceedings of International Conference on Agile Software Development*, 2013, pp. 61-75.
- [8]. D. Donmez, G. Grote, Two sides of the same coin: How agile software development teams approach uncertainty as threats and opportunities, *Information* and Software Technology, Vol. 93, 2018, pp. 94-111.
- [9]. A. Fellah, Embracing residuality theory in software architecture to address uncertainty: Key challenges and strategies, in *Proceedings of the International Conference on Software Engineering and Data Engineering*, 2024, pp. 31-34.
- [10]. R. Hoda, J. Noble, A grounded theory of agile transitions in practice, in *Proceedings of the IEEE/ACM 39th International Conference on Software Engineering (ICSE'17)*, Buenos Aires, Argentina, 2017, pp. 141-151.
- [11]. M. Merkow, Secure, Resilient, and Agile Software Development, *CRC Press*, 2019.
- [12]. E. Normand, Residuality Theory: Good Idea, Bad Name, https://ericnormand.substack.com/p/residualitytheory
- [13]. B. M. O'Reilly, An introduction to residuality theory: Software design heuristics for complex systems, *Procedia Computer Science*, Vol. 170, 2020, pp. 875-880.
- [14]. N. N. Taleb, Antifragile: Things That Gain From Disorder, *Random House*, 2012.

- [15]. M. Waterman, Agility, risk, and uncertainty, part 2: How risk impacts agile architecture, *IEEE Software*, Vol. 35, Issue 3, 2018, pp.18-19.
- [16]. K. Schwaber, Scrum development process, in Proceedings of the OOPSLA Conference, 1995, pp. 117-127.
- [17]. K. Schwaber, J. Sutherland, The Scrum Guide, *Scrum.org*, 2011, pp. 1-38.
- [18]. S. Shore, S. Warden, The Art of Agile Development, *O'Reilly Media*, 2012.
- [19]. M. Beedle, et al., Agile Manifesto, http://www.agilemanifesto.org

(043)

The Role of Code Readability in Large Language Model Code Summarization

<u>B. Szalontai</u>¹, G. Szalay¹, T. Márton¹, A. Sike¹, P. Mátray², M. I. Nagy², B. Pintér¹ and T. Gregorics¹

¹ Eötvös Loránd University, Faculty of Informatics, Department of Software Technology and Methodology, 1/C Pázmány Péter sétány, 1117, Budapest, Hungary ² Ericsson, 11. Magyar Tudósok Körútja, 1117, Budapest, Hungary Er meilt hulm00@infolta.hu

E-mail: bukp00@inf.elte.hu

Summary: Large Language Models (LLMs) have demonstrated good performance in various software engineering tasks, including code summarization - explaining what a code snippet does. In this paper, we argue that the *readability* of the input code is crucial to how well the LLM will be able to explain it: a readable code has a higher chance of being correctly explained. We show that metrics that estimate code readability correlate with the ability of LLMs to explain what the code does and can also be used to predict how well the LLM will do. We analyse the human-interpretable readability features used for the prediction to characterize the code snippets that can be explained well. We demonstrate the connection between code readability and explainability on one of the most widely acknowledged code explanation benchmarks, HumanEvalExplain, and across six different open LLMs.

Keywords: Large language models, Code explanation, Code summarization, Code readability.

1. Introduction

Recently, large language models (LLMs) have achieved remarkable results in software engineeringrelated tasks [1-7], with some specifically designed for this domain [8-15]. While code generation (e.g. HumanEval [16], APPS [17], MBPP [18]) and bug fixing (e.g. QuixBugs [19], HumanEvalFix [20]) are among the most popular software engineering benchmark tasks for LLMs, there has also been some work on code summarization, where the model has to explain what the code does (e.g. CodeXGLUE [21], HumanEvalExplain [20]). It is worth noting that results on such benchmarks are not usually used for comparing or ranking LLMs.

While many LLMs exhibit remarkable results in software engineering tasks, what exactly happens inside them, and whether they truly comprehend their inputs, remains an open question. It has been hypothesized for example, that certain code properties are not encoded by LLMs, which might cause generating erroneous code [22]. Others have explored the limitations of code analysis performed by LLMs, pointing to potential problems with their understanding of dynamic behaviours [23]. Furthermore, it has been suggested that as code becomes more complex (i.e. using nested constructs, complicated loops, or nontrivial operators), it challenges the reasoning ability of LLMs [24].

An interesting and less examined question is looking at the problem from the other angle: instead of trying to characterize how and which LLMs can solve the problem of code summarization, we can try to characterize the kind of code that is explainable (summarizable) by LLMs and the kind of code that isn't. To investigate this problem, we turn to the concept of code readability [25, 26]. The readability of code refers to the clarity and ease with which a human can interpret its meaning and logic. Readability can be influenced by factors such as code structure, identifier names, or code complexity, but it can be subjective and can vary from developer to developer. Estimating source code readability has long been an interesting field of research with many methods proposed [25-34].

We hypothesize that a connection exists between code readability and explainability. This is evidently the case for humans: a more readable code is generally easier to comprehend and explain. We suggest that this principle applies similarly to the code explanation capability of LLMs: LLMs should explain readable code snippets better than less readable ones.

In this paper, we try to characterize the kind of programs that can be summarized well (are explainable) by LLMs using code readability measures and features. First, we examine the correlation between code readability and explainability using the readability metric proposed by Posnett et al. [26]. Next, we use the human-interpretable features of two readability metrics [25, 26] to build classifiers that predict how well LLMs will be able to summarize a code snippet. We obtain our results on the most widely code explanation benchmark, used HumanEvalExplain. The results are consistent across six currently popular open LLMs.

Our contributions are as follows:

- We show that significant correlation exists between code readability and the ability of LLMs to summarize code;
- We build human-interpretable classifiers that can predict whether a code will be correctly summarized by an LLM;

• We analyse the obtained classifiers and extract the features that determine the explainability of code snippets.

2. Related Work

Now we turn to works related to investigating the code comprehension abilities of LLMs. Also, as we believe that program readability correlates with whether the program can be well summarized by LLMs, we provide an overview of proposed code readability metrics.

2.1. Code Comprehension Ability of LLMs

Anand et al. [22] investigated attention maps and hidden representations of LLMs. They conducted their study on 3000 randomly sampled Python programs from the CodeSearchNet dataset [35]. The used models include encoder-only (CodeBERT [36] GraphCodeBERT [37]), encoder-decoder (CodeT5 [38], PLBART [39], CodeT5+ [40]) and decoder-only (CodeGen [41]) models. Their analysis revealed that LLMs struggle to encode relations between syntactic and identifier tokens, restricting their ability to comprehend program flow and logic. They also observed that fine-tuned and larger models encode these relations more poorly compared to smaller pretrained models. Their findings show that fine-tuned and larger models rely on shortcut learning and memorize code instead of code comprehension.

Ma et al. [23] categorized LLM code analysis capabilities into syntax understanding, static behaviour understanding, and dynamic behaviour understanding. Their study showed that while LLMs handle syntax and static analysis relatively well, they struggle with dynamic behaviours, such as reasoning about test flakiness and equivalent mutant detection.

Sun et al. [32] examined LLMs' performance in generating code summaries across different programming paradigms. They found that LLMs perform worse on logic programming languages (e.g., Prolog) compared to procedural and object-oriented languages, suggesting that underlying language structure affects explainability.

Liu et al. [24] introduced CodeMind to evaluate LLMs' reasoning abilities, showing that performance declines as code complexity increases. LLMs particularly struggle with nested constructs, complex conditions, and API invocations, indicating that high complexity hinders their ability to reason about execution.

This suggests that as code complexity increases, LLMs face significant challenges in reasoning about code execution.

2.2. Estimating Code Readability

There have been many attempts to estimate code readability, some of which purely rely on syntactic or

visual code features, while others use deep learningbased approaches.

Buse et al. train readability classifiers [25] based on judgments of 120 human annotators, who have annotated 100 code snippets. Based on the feedback of annotators, they suggest 25 local code features that correlate with code readability, such as length of identifiers or number of blank lines. Posnett et al. propose a simpler model of code readability [26], improving on the metric of Buse et al. They focus on three structural features: number of lines, entropy, and Halstead's Volume. The readability score is determined by combining these features. Compared to Buse et al.'s data [25], this model more closely correlates with human judgments.

Dorn et al. present a generalizable formal model of software readability [28], based on a study of 5000 participants. Their approach analyses visual, spatial, and linguistic features, such as structural patterns, code block sizes, and identifier content, across Java, Python, and CUDA. Human annotators evaluate readability on programs up to 50 lines long. Similarly, Scalabrino et al. [29,30] emphasize the importance of both syntactic and textual aspects of source code readability. They propose features like comment and identifier consistency, textual coherence, and the number of meanings and concepts. Their model is validated on Java, Python, and CUDA as well.

Mi et al. propose IncepCRM [31], a deep learning-based model for classifying code as readable or unreadable using convolutional operators and auxiliary annotator inputs. Later, they introduce a hybrid neural network [32] combining BERT, CNN, and BiLSTM to extract readability features from RGB matrices (visual), token sequences (semantic), and character matrices (structural). Hu et al. [33] extend this line of research to code maintainability with DeepM, a recurrent model using LSTM, tree-LSTM, and attention mechanism. They categorize programs from GitHub as high or low quality based on metrics like stars and contributor activity. The categorized programs are then used as training data for DeepM.

3. Method

In this Section, we describe the dataset, the readability metrics, the LLMs, how we measure correlation between readability and explainability, and how we predict explainability from readability using human-interpretable features.

3.1. Dataset

We use the HumanEvalExplain benchmark dataset [20], which is probably the most acknowledged benchmark for code summarization by LLMs. It is one of the benchmarks in HumanEvalPack, which contains variants of the HumanEval code generation benchmark.

The benchmark contains 164 programs. Each program is relatively short and designed to solve a

commonly encountered programming task, such as determining if a number is prime.

Models are evaluated as follows. The evaluated model first generates an explanation of the code, then tries to reconstruct the original code based on its explanation. Each reconstructed code is validated using test cases. The outcome of this validation process is binary for each input: a pass@1 score of 1 indicates that the explanation resulted in a successful code generation, while a score of 0 denotes failure due to inaccurate explanation or incorrect code reconstruction.

3.2. Code Readability Metrics

For evaluating code readability, we implement the readability metrics proposed by Buse et al. and Posnett et al. [25, 26]. Both metrics estimate the readability of a code snippet using extracted features. Buse et al. determined 25 such features which they deemed important for code readability. Most of these features are calculated either as an average value per line or maximum value in any one line. We obtain the values for these features from a code snippet either by using a tokenizer and processing the code as a set of tokens or extract the features directly from the code, in which case it is processed as a string. In addition to features such as line length in characters, number of keywords or number of loops, there are two features which calculate the occurrences of a character or an identifier and return the most frequent one's frequency.

In Python, code can be indented using either spaces or tab characters. In our implementation, we count each space as one unit and each tab as four units.

Posnett et al. propose a simpler model of code readability, relying on three structural code features: number of lines, code entropy, and Halstead's Volume metric.

3.3. Large Language Models

We use six open and instruction-tuned models for our evaluations, with model sizes ranging from 1.5B to 70B parameters. The evaluated models are Qwen2.5-Coder (1.5B),Llama-3.2 (3B), DeepSeekCoder (6.7B), Codestral (22B), Mixtral (8×7B), and Llama-3.3 (70B). Half of these models are general-purpose LLMs, while the other half are specialized LLMs designed to function as coding assistants. We ran the HumanEvalExplain benchmark on these models with fp16 precision in a pass@1 setting with greedy decoding (no sampling) for generating both the explanation and code. The results can be seen in Table 2.

3.4. Correlation between Readability and explainability

In order to measure correlation between code readability and explainability, we first divide the programs of the benchmark dataset into two groups based on these scores. Evidently, this grouping will be different for each evaluated model.

We use the readability model proposed by Posnett et al. The readability score is computed for each program, so the distributions of scores between the two groups (successful and failed explanation) can be compared. We compare these using boxplots, pointbiserial correlation, and Mann-Whitney U-tests. These comparisons are repeated for each LLM.

3.5. Predicting Explainability from Code Readability Features

Estimators of code readability scores are usually classifiers or regressors working with certain sets of features extracted from code. The readability score could be, for example, the probability assigned by logistic regression [25]. Instead of using the entire code readability metric and measuring its correlation with the explainability of code, we can directly use the readability features to predict how well LLMs can explain a code snippet.

We are using two sets of features: in addition to the features of the Posnett readability metric, we're also using the features of Buse et al. [25]. Although the readability metric of Posnett et al. is more recent and achieves better results, the features of Buse et al. are more numerous and interesting from an interpretability standpoint. We compute these features for each program in the HumanEvalExplain dataset and use them in our classifiers. The targets are the explainability scores (0 or 1 for each program).

As the explainability of a code snippet is binary in benchmark datasets (it was either correctly explained or not), our prediction task will be binary classification. We have chosen two classifiers where the importance of individual features to the classification decision is easy to determine: logistic regression and random forest. We can also contrast the simple model (logistic regression), and the complex one (random forest).

To compensate for the small size of the dataset, we use K-fold cross-validation with K = 100. Since we consider six LLMs, two sets of readability features, and 100 folds, we train $6 \cdot 2 \cdot 100 = 1200$ classifiers of both logistic regression and random forest.

4. Results

We first present our results about the correlation between readability and code explainability. Then we train classifiers to predict explainability directly from the features of code readability and perform feature analysis on the trained classifiers. We investigate results across the six open LLMs.

4.1. Correlation between Readability and Explainability

We split the dataset into two groups according to whether the summarization was successful or not, then

we analyse the readability scores of these groups. So, we obtain two distributions of readability scores (successful, failed) and compare them. Fig. 1 shows the boxplots of these distributions across the six LLMs. We also include the exact values of the medians (and the means) in Table 1.



Fig. 1. Boxplots that visualize the readability scores of programs in the HumanEvalExplain dataset. The programs are categorized into sets of programs with successful and failed summarizations.

To get a clearer and more detailed picture, we perform kernel density estimation (KDE). This offers a smoothened visualization of the probability distributions of readability scores of the two groups (Fig. 2).



Fig. 2. Kernel density estimation (KDE) plots depicting the distribution of readability scores for programs classified as having successful and failed summarization, for each evaluated LLM.

Beyond visualization, we use statistical testing to evaluate the relationship between readability and explainability. We conduct Mann-Whitney U-tests to compare the distributions of readability scores between successful and failed summarization groups. The resulting p-values can be seen in Table 2.

To quantify the relationship further, we calculate the point-biserial correlation coefficient between readability scores and the group (successful or failed summarization) for each LLM. The correlation coefficients and p-values can be found in Table 2.

4.2. Predicting Explainability from Readability Features

A step beyond measuring correlation is trying to predict if the LLM will be able to summarize a code snippet from the readability of the snippet. In this Section, we train classifiers on readability features to answer this question. We would also like to determine the features that decide explainability.

Keeping these goals in mind, we chose two classifiers: random forest and logistic regression with L1 regularization (Lasso). The task is to predict whether the code snippet will be summarized correctly by the LLM from the readability features of the code snippet (Table 1). Fig. 3 shows the accuracy across the six LLMs and two sets of readability features. The boxplots show the performance across the 100 folds of cross-validation. Each tick on the x-axis corresponds to a different LLM.

Table 1. The median and mean readability scores of programs grouped by summarization outcomes (successful (S) or failed (F)). Results are rounded to 2 decimals.

LLM	Median (S)	Median (F)	Mean (S)	Mean (F)
Llama-3.2 (3B)	0.58	0.36	0.54	0.42
Mixtral (8×7B)	0.66	0.25	0.56	0.35
Llama-3.3 (70B)	0.60	0.21	0.53	0.34
Qwen2.5-Coder (1.5B)	0.60	0.22	0.52	0.31
DeepSeekCoder (6.7B)	0.61	0.06	0.54	0.26
Codestral (22B)	0.60	0.06	0.52	0.24

 Table 2. Mann-Whitney U-tests and point-biserial

 correlations. Results on the HumanEvalExplain benchmark

 is also displayed for comparison.

LLM	U-test p-value	r_pb	p_corr	Results (%)
Llama- 3.2 (3B)	$5.14 \cdot 10^{-2}$	0.11	$1.62 \cdot 10^{-1}$	12.80
$\begin{array}{c} \text{Mixtral} \\ (8 \times 7\text{B}) \end{array}$	9.02 · 10 ⁻⁵	0.29	$2.10 \cdot 10^{-4}$	40.85
Llama- 3.3 (70B)	4.58 · 10 ⁻⁴	0.27	7.50 · 10 ⁻⁴	50.00
Qwen2.5 -Coder (1.5B)	2.21 · 10 ⁻⁵	0.30	1.07 · 10 ⁻⁴	57.32
DeepSee kCoder (6.7B)	2.91 · 10 ⁻⁷	0.37	8.28 · 10 ⁻⁷	62.20
Codestral (22B)	9.32 · 10 ⁻⁷	0.36	$1.72 \cdot 10^{-6}$	67.68

7th International Conference on Advances in Signal Processing and Artificial Intelligence (ASPAI' 2025), 8-10 April 2025, Innsbruck, Austria



Fig. 3. Explainability results of random forest and logistic regression classifiers of different LLMs using both Buse and Posnett readability features.

4.2.1. Feature Importance Analysis

We examine the importance of the features in the trained classifiers. For the L1 logistic regression, we calculate the number of folds where the feature is active divided by the total number of folds, so we get the percentage of folds where the feature is active. We calculate this separately for each LLM, so we can compare which features are typically active for which LLM (Fig. 4). Features with consistently high utilization (such as the line length for Buse's metric or Halstead's Volume for Posnett's metric) play a larger role in classification.

The bar plot of Fig. 5 visualizes the mean and standard error of feature importance determined by the random forest classifier, macro-averaged across LLMs and folds. Features are ordered by their importance, and error bars indicate the standard error across LLMs. The results highlight which features contribute most significantly to the classification.

5. Discussion

The first question we asked was whether code readability scores correlate with the success of summarization. The boxplots (Fig. 1) and KDE plots (Fig. 2) show that there is significant difference between the readability score distributions of code snippets whose summarization was successful and snippets whose summarization failed. Table 2 shows the exact medians and means of these distributions. On average, the programs where summarization was successful have $4.99 \times$ higher medians and $1.72 \times$ higher means. Correlation is strongest for the

DeepSeekCoder and Codestral models, with around a $10\times$ higher median and more than $2\times$ higher mean readability for the successfully summarized group. The connection was further strengthened by the p-values obtained in the statistical tests (Table 2).



Fig. 4. Feature utilization of the L1 logistic regression. Each cell in the heatmap indicates the percentage a feature was active across folds.



Fig. 5. Mean and Standard Error of Feature Importance using the random forest classifier.

The connection between readability and summarizability seems to be general as it is strong across all language models, except for Llama-3.2 (3B), where the p-values are not significant. We believe this is because this model was especially weak in code summarization: as Table 2 shows, it achieved only 12.8 % in code summarization, where other models achieved 40.9 % to 67.7 %.

The second question was whether it is possible to predict whether an LLM will summarize code successfully using explainable readability features. To answer this, we evaluated random forest and Lasso logistic regression classifiers for each LLM, using a 100-fold cross validation. Based on Fig. 3, which shows the accuracy of this prediction across the folds, the prediction is possible.

We have chosen these classifiers and features since their results are explainable, and we were curious which features are important for summarizability. The heatmaps in Fig. 4 visualize the features selected by logistic regression with L1 (Lasso) regularization across the 100 folds. Mostly the same features are selected across all LLMs.

Among the heatmap visualizing features proposed by Buse et al., the most significant are the average and maximum line length, the count of the most common identifier and character, and the maximum identifier length. From the three features used by Posnett et al., Hastead Volume and Entropy are consistently selected across LLMs, while the number of lines matters only for some.

The same Posnett features are selected by random forest as by Lasso (Fig. 5). The situation is somewhat different for the features of Buse et al. Some of the same features are selected, like the count of the most common identifier and character. The others selected by Lasso (line length maximum and average, identifier length maximum) are also significant for random forest, but there are also some other features just as or more significant, like average number of assignments, and a host of other features that average.

6. Conclusion

In this work, we demonstrated that there is a strong connection between the readability of code and how well LLMs can explain (summarize) it. First, we demonstrated the correlation between code readability and explainability, then we built classifiers based on readability features to predict whether LLMs can summarize the code snippet correctly.

Lastly, we analysed which readability features are important for code summarizability. These include the average and maximum line length, as well as the frequency of single characters and identifiers. Out of the three features of Posnett et al., the feature analysis has pointed out Halstead's Volume and entropy.

Our results were consistent across six open LLMs on the most widely used code summarization benchmark, HumanEvalExplain.

In future work, we aim to examine the connection between code readability and explainability in more general settings: we would like to extend our experiments to a larger-scale benchmark with longer and more complex code snippets, and we would like to make measurements on different programming

References

[1]. OpenAI, GPT-4 technical report, *arXiv preprint*, Mar. 2023, arXiv:2303.08774.

- [2]. Gemini: A family of highly capable multimodal models, https://arxiv.org/abs/2312.11805
- [3]. H. Touvron, et al., Llama 2: open foundation and fine-tuned chat models, https://arxiv.org/abs/ 2307.09288
- [4]. A. Dubey, et al., The Llama 3 herd of models, https://arxiv.org/abs/2407.21783
- [5]. J. Bai, et al., Qwen technical report, https://arxiv.org/ abs/2309.16609
- [6]. A. Q. Jiang, et al., Mixtral of experts, https://arxiv.org/ abs/2401.04088
- [7]. DeepSeek-V2: A strong, economical, and efficient mixture-of-experts language model, https://arxiv.org/ abs/2405.04434
- [8]. Q. Zheng, et al., CodeGeeX: A pre-trained model for code generation with multilingual benchmarking on HumanEval-X, in Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Aug. 2023, pp. 5673-5684.
- [9]. R. Li, *et al.*, StarCoder: may the source be with you!, *arXiv preprint*, May 2023, arXiv:2305.06161.
- [10]. Z. Luo, et al., WizardCoder: empowering code large language models with evol-instruct, https://arxiv.org/ abs/2306.08568
- [11]. B. Roziere, et al., Code Llama: open foundation models for code, arXiv preprint, Aug. 2023, arXiv:2308.12950.
- [12]. D. Guo, et al., DeepSeek-Coder: when the large language model meets programming -- the rise of code intelligence, https://arxiv.org/abs/2401.14196
- [13]. Y. Wei, Z. Wang, J. Liu, Y. Ding, L. Zhang, Magicoder: source code is all you need, https://arxiv.org/abs/2312.02120
- [14]. Z. Yu, et al., WaveCoder: widespread and versatile enhancement for code large language models by instruction tuning, https://arxiv.org/abs/2312.14187
- [15]. B. Hui, et al., Qwen2.5-Coder technical report, https://arxiv.org/abs/2409.12186
- [16]. M. Chen, et al., Evaluating large language models trained on code, https://arxiv.org/abs/2107.03374
- [17]. D. Hendrycks, et al., Measuring coding challenge competence with APPS, https://arxiv.org/abs/ 2105.09938
- [18]. J. Austin, et al., Program synthesis with large language models, https://arxiv.org/abs/2108.07732
- [19]. D. Lin, J. Koppel, A. Chen, A. Solar-Lezama, QuixBugs: a multi-lingual program repair benchmark set based on the Quixey challenge, in *Proceedings of* the ACM SIGPLAN International Conference on Systems, Programming, Languages, and Applications: Software for Humanity, Oct. 2017.
- [20]. N. Muennighoff, et al., OctoPack: Instruction tuning code large language models, https://arxiv.org/abs/ 2308.07124
- [21]. S. Lu, *et al.*, CodeXGLUE: a machine learning benchmark dataset for code understanding and generation, https://arxiv.org/abs/2102.04664
- [22]. A. Anand, S. Verma, K. Narasimhan, M. Mezini, A critical study of what code-LLMs (do not) learn, https://arxiv.org/abs/2406.11930
- [23]. W. Ma, et al., The scope of ChatGPT in software engineering: a thorough investigation, https://arxiv.org/ abs/2305.12138
- [24]. C. Liu, S. D. Zhang, A. R. Ibrahimzada, R. Jabbarvand, CodeMind: a framework to challenge large language models for code reasoning, https://arxiv.org/abs/ 2402.09664

- [25]. R. P. L. Buse, W. Weimer, A metric for software readability, in *Proceedings of the International Symposium on Software Testing and Analysis* (ISSTA'08), 2008, pp. 121-130.
- [26]. D. Posnett, A. Hindle, P. Devanbu, A simpler model of software readability, in *Proceedings of the 8th Working Conference on Mining Software Repositories*, May 2011.
- [27]. R. P. L. Buse, W. R. Weimer, Learning a metric for code readability, *IEEE Transactions on Software Engineering*, Vol. 36, Issue 4, Jul. 2010, pp. 546–558.
- [28]. J. Dorn, A general software readability model, https://web.eecs.umich.edu/~weimerw/students/dornmcs-pres.pdf
- [29]. S. Scalabrino, M. Linares-Vasquez, D. Poshyvanyk, R. Oliveto, Improving code readability models with textual features, in *Proceedings of the IEEE 24th International Conference on Program Comprehension* (*ICPC'16*), Austin, TX, USA, 2016, pp. 1-10.
- [30]. S. Scalabrino, M. Linares-Vásquez, R. Oliveto, D. Poshyvanyk, A comprehensive model for code readability, *Journal of Software: Evolution and Process*, Vol. 30, Issue 6, Jun. 2018, e1958.
- [31]. Q. Mi, J. Keung, Y. Xiao, S. Mensah, X. Mei, An inception architecture-based model for improving code readability classification, in *Proceedings of the 22nd International Conference on Evaluation and Assessment in Software Engineering (EASE'18)*, 2018, pp. 139-144.
- [32]. Q. Mi, Y. Hao, L. Ou, W. Ma, Towards using visual, semantic and structural features to improve code readability classification, *Journal of Systems and Software*, Vol. 193, Jul. 2022, pp. 111454–111454.
- [33]. Y. Hu, H. Jiang, Z. Hu, Measuring code maintainability with deep neural networks, *Frontiers of Computer Science*, Vol. 17, Issue 6, Jan. 2023.
- [34]. K. Park, *et al.*, An eye tracking study assessing source code readability rules for program comprehension,

Empirical Software Engineering, Vol. 29, Issue 6, Oct. 2024, 160.

- [35]. H. Husain, H.-H. Wu, T. Gazit, M. Allamanis, M. Brockschmidt, CodeSearchNet challenge: evaluating the state of semantic code search, https://arxiv.org/abs/1909.09436
- [36]. Z. Feng, et al., CodeBERT: a pre-trained model for programming and natural languages, in Findings of the Association for Computational Linguistics: EMNLP 2020, Association for Computational Linguistics, Feb. 2020, pp. 1536-1547.
- [37]. D. Guo, et al., GraphCodeBERT: Pre-training Code Representations with Data Flow, https://openreview.net/forum?id = jL oC4ez43PZ
- [38]. Y. Wang, W. Wang, S. Joty, S. C. H. Hoi, CodeT5: Identifier-aware Unified Pre-trained Encoder-Decoder Models for Code Understanding and Generation, in Proceedings of the Conference on Empirical Methods in Natural Language Processing, Jan. 2021, pp. 8696-8708.
- [39]. W. Uddin Ahmad, S. Chakraborty, B. Ray, K.-W. Chang, Unified Pre-training for Program Understanding and Generation, in *Proceedings of the Conference of the North American Chapter of the Association for Computational*, Jun. 2021, pp. 2655-2668.
- [40]. Y. Wang, H. Le, A. D. Gotmare, N. D. Q. Bui, J. Li, S. C. H. Hoi, CodeT5+: open code large language models for code understanding and generation, https://arxiv.org/abs/2305.07922
- [41]. E. Nijkamp, et al., CodeGen: an open large language model for code with multi-turn program synthesis, https://arxiv.org/abs/2203.13474
- [42]. W. Sun, et al., Source code summarization in the era of large language models, https://arxiv.org/abs/ 2407.07959

(044)

Traffic Predictions Using Graph Neural Networks on Real-time Observations

Joachim Hansen¹, <u>Donglin Liu</u>² and Alexandros Sopasakis²

¹ Department of Physics, Lund University, 22362 Lund, Sweden
² Department of Mathematics, Lund University, 22362 Lund, Sweden E-mail: donglin.liu@math.lth.se

Summary: We gather real-time traffic data from Trafikverket's extensive camera network in Gothenburg and harness cutting-edge graph neural networks (GNNs) to generate precise predictions of road traffic density for the next hour. Each camera is treated as a graph node with edges representing road connectivity and spatial proximity. Vehicle detection is performed using YOLOv5, which generates accurate density metrics while filtering out background noise from shadows and reflections. We observed that increasing the training period from 3 days to 14 days generally leads to improved forecasting accuracy, as indicated by a reduction in the mean absolute percentage error (MAPE). For example, under the GWNET model, the MAPE decreased from 53.13 % with 3-day training to 47.28 % with 14-day training, demonstrating that longer training periods enable the model to better capture the underlying spatiotemporal dynamics. These results demonstrate the robustness of GNNs in traffic forecasting and underscore that abundant data is essential – not only for capturing rush-hour fluctuations but also for investigating daily and weekly patterns in future studies.

Keywords: Graph neural network, Spatial dependence, Traffic prediction.

1. Introduction

Graph neural networks (GNNs) have emerged as a powerful framework for modeling complex relationships in data structured as graphs. Unlike traditional neural networks, GNNs operate directly on graph-structured data, making them particularly well-suited for tasks where spatial relationships are important. In the realm of traffic prediction, where understanding not only temporal structure but also spatial dependencies among road segments, GNNs offer significant advantages [1].

At their core, GNNs leverage the inherent structure of graphs to capture spatial dependencies and patterns. In traffic prediction tasks, roads and intersections can be naturally represented as nodes in a graph, with edges denoting connections or proximity between them. By incorporating this spatial information, GNNs can effectively model how traffic conditions propagate through a road network.

One key strength of GNNs lies in their ability to aggregate information from neighboring nodes in the graph. Through iterative message passing schemes, GNNs accumulate and refine information from neighboring nodes, allowing them to capture complex spatial dependencies and patterns. This mechanism enables GNNs to learn representations that encode not only local characteristics of individual road segments but also global properties of the entire road network [2]. Moreover, GNNs can dynamically adapt to changing traffic conditions. By incorporating temporal information alongside spatial features, GNNs can capture how traffic patterns evolve over time. This temporal awareness allows GNNs to make accurate predictions even in dynamic traffic environments, where conditions may change rapidly [1-4]. Furthermore, GNNs facilitate learning, where

additional data sources such as weather conditions and traffic cameras. By jointly modeling spatial, temporal, and additional contextual information, GNNs offer a holistic approach to traffic prediction that outperforms traditional methods.

2. Previous Work

At their core, GNNs leverage the inherent structure of graphs to capture spatial dependencies and patterns. In traffic prediction tasks, roads and intersections can be naturally represented as nodes in a graph, with edges denoting connections or proximity between them. By incorporating this spatial information, GNNs can effectively model how traffic conditions propagate through a road network.

Recent advances in traffic forecasting have explored a variety of deep learning architectures to capture the complex spatio-temporal dynamics inherent in urban traffic. Li et al. [5] introduced the diffusion convolutional Recurrent Neural Network (DCRNN), which models traffic flow as a diffusion process over graphs to effectively capture continuous spatial-temporal dependencies. However, despite its strengths, DCRNN does not explicitly address abrupt phase transitions and requires large, diverse datasets for optimal performance. Similarly, Zhang et al. [6] developed a deep spatio-temporal residual network for citywide crowd flow prediction, integrating recent, periodic, and trend components. Although effective for aggregated flows, this approach struggles to differentiate between distinct traffic phases.

In the context of Gothenburg, earlier studies [7] have primarily focused on the temporal dynamics of traffic flow using data from individual cameras, often employing Long Short-Term Memory (LSTM)

networks [8] – one of the three neural architectures we compare in our study. Recognizing that traffic systems are not solely temporal but also heavily influenced by spatial topology, our work extends these frameworks by integrating data from multiple cameras. In our model, each camera is represented as a node defined by its GPS coordinates and viewing direction, while edges are formed based on the physical distances between cameras with direct road connections. This graph-based formulation yields a simulated traffic network that more accurately reflects the actual layout of Gothenburg's roads and intersections, providing a more comprehensive approach to analyzing urban traffic dynamics.

2. Traffic Data and Processing

The dataset originates from traffic cameras installed at various intersections throughout Gothenburg by Trafikverket. Each camera is equipped with GPS coordinates that pinpoint its precise location and define its field of view. We model each camera as a node within a directed graph where edges represent the road distances between connected cameras. Over 27 consecutive days (from 5 am to 9 pm), minute-by-minute images were collected from 57 cameras. For this study, we extracted subsets spanning 3, 7, and 14 consecutive days to assess the impact of training-set size on predictive accuracy, with each larger datasets fully incorporating the data from the smaller one. This integrated approach yields a simulated traffic network that closely approximates the actual layout of roads and intersections.

Accurately converting images into traffic density measurements requires precisely identifying and counting vehicles under varying weather and lighting conditions. We use YOLOv5 [9] for vehicle detection, which offers a robust alternative to traditional counting methods that are sensitive to lighting variations [7] (see Fig. 1). Once vehicles are detected, we calculate pixel counts as a fraction of the total road area to derive density estimates. These values are then averaged over five-minute intervals to smooth out transient fluctuations. The resulting density measurements, along with the constructed spatial graph, form the input to a Graph Neural Network (GNN) that captures both the temporal dynamics and spatial relationships within the data. It is worth noting that because the entire image frame is used to define the region of interest not just the road - density values can occasionally exceed one, although this is rare in practice. Moreover, transforming the raw data into a neural network-ready format involves several critical steps. In the subsequent section, we describe these processes in detail.



Fig. 1. Artifacts and noise from shadows and reflections from traditional, filtering, algorithms for density estimation (left and middle) versus Yolov5 method (right).

3.1. Image Pre-processing

Initially, raw images collected from various cameras across Gothenburg are enriched with corresponding GPS coordinates. These coordinates are then used to build a graph that encapsulates the spatial relationships among the cameras, ensuring that the input to the neural network effectively represents road occupancy over short temporal windows.

Traditional filtering methods [7] that combine edge detection with Gaussian blurring may lack precision in delineating vehicle boundaries. As shown in Fig. 1, these techniques often yield inaccurate vehicle density estimates throughout the day, primarily due to sunlight reflections on the road being misclassified as vehicles. To overcome these limitations, we employ the YOLOv5 model, pre-trained on the COCO dataset [10]. This approach reliably detects vehicles by generating bounding boxes across varying lighting conditions, see Fig. 1.

3.2. Graph Construction

The raw GPS data consist of latitude and longitude coordinates, along with a bearing that indicates each camera's field of view. To optimize the construction of our data graph, we represent camera locations as nodes and establish directed edges between them, forming a digraph (a directed graph). Self-connections between nodes are not permitted.

Since cameras monitor roads in specific traffic directions, it is essential to establish node connections based on whether traffic flows in the direction the camera is facing. For instance, consider Camera X and Camera Y: if traffic flows in the direction that Camera X is oriented – meaning Camera Y lies within its forward-facing view cone – a connection is established from X to Y. Conversely, if traffic flows opposite to the direction of Camera X's view, no connection is formed from X to Y.

We manually checked each camera to determine the road direction and adjusted their respective masks (the green region shown in Fig. 1) accordingly. On bidirectional roads, for instance, where a single camera may capture traffic moving in both directions, we employ directional masks to separate and process the information for each direction individually. This approach ensures that node connections accurately reflect the observed traffic dynamics, with connections only formed when traffic flows in the camera's designated direction.

Each camera's GPS coordinates and viewing direction define a view cone. If another camera falls within a threshold distance R and an angular tolerance $\Delta \varphi$, an edge is formed – potentially bidirectional, depending on observed traffic flow. This process yields the adjacency matrix A that captures spatial relationships. Down-sampling to 5-minute intervals produces a time series per day for each of the 57 nodes. Following density estimation, the image datasets are then transformed into a tensor with dimensions (B,F,S,V), where B denotes the batch size, F comprises features such as estimated density and temporal encodings, S represents the time slice sample, and V indicates the number of cameras. The final graph is shown in Fig. 2.



Fig. 2. One representation of the data graph based on the GPS camera locations.

3.3. Time-dependent Laplacian

In spectral graph theory [11, 12], a graph's structure is often characterized by its Laplacian matrix, denoted by L. The Laplacian matrix is defined as,

$$L = D - A,\tag{1}$$

where A is the adjacency matrix that represents the connections between nodes, and D is a diagonal matrix whose entries represent the node degrees. A node degree is a measure of a node's connectivity

within the graph. For node \$i\$, the degree is given by $D_{ii} = \sum_{j} A_{ij}$, which sums the weights of all edges connected to node *i*. In unweighted graphs, this sum equals the number of connections (or edges) that node *i* has. In weighted graphs, it represents the total weight of the connections. This concept is critical, as it quantifies the importance or centrality of a node within the network.

While in many applications the Laplacian matrix remains static upon its creation, in the context of traffic dynamics, this assumption is only conditionally valid. Particularly during peak traffic periods such as morning and afternoon rush hour, there tends to be a directional preference in the flow of traffic. This directional bias arises from commuters traveling to and from work, contributing to the temporal variability of traffic patterns. Consequently, there arises a need to model this time-varying behavior by transforming the Laplacian matrix as,

$$L \to L(t) \tag{2}$$

Several methods have been proposed to effect this transformation. One approach employs neural networks to predict and generate the edge strengths of the adjacency matrix, thereby capturing the evolving dynamics of traffic flow. However, traffic dynamics exhibit distinct phases [13, 14] – triggered by density fluctuations – and are subject to rapid shifts. Consequently, training neural networks to accurately capture such multi-phase dynamics requires a substantial amount of data to learn optimal transformations of the Laplacian matrix. Early attempts to apply deep neural networks for this purpose have encountered challenges [5, 6], particularly due to limited dataset sizes, which hamper effective learning.

To address these challenges, we propose a dynamic modification of the adjacency matrix $A \rightarrow A(t)$, defined as

$$A(t) = (1 - \omega)A_0 + \omega A'(t), \qquad (3)$$

where A_0 denotes the static adjacency matrix constructed from GPS data, and A'(t) is a time-varying adjustment that reflects evolving traffic patterns. The parameter $\omega \in [0,1]$ governs the trade-off between the inherent static connectivity of the road network and its dynamic, temporal variations. Based on preliminary experiments, an equal weighting $(\omega = 0.5)$ provided a robust balance, ensuring that neither the stable structural features nor the transient fluctuations dominate traffic the learned representation. In our formulation, A_0 is our initial adjacency matrix estimated from GPS data as explained in Section 3.2, and A'(t) is re-evaluated at each time iteration via a linear transformation of the updated input matrix as follows,

$$A'(t)_{57\times57} = Q_{57\times2}M(t)_{2\times12}P_{12\times57},$$
(4)

where Q and P are transformation matrices that map an intermediate, lower-dimensional representation back into the full node space. In our framework, we set $Q \in \Re^{57\times2}$, $M \in \Re^{2\times12}$, and $P \in \Re^{12\times57}$ so that the product $QM(t)P \in \Re^{57\times57}$ matches the dimensions of the underlying network which is described in Section 3.4. This design ensures that the learned transformations are directly compatible with the network structure. By leveraging the geometric structure of the GPS to map the intricate connection of cameras, combined with the temporal information embedded within the network and the Laplacian, this approach establishes a robust foundation for modeling traffic density.

3.4. Architectures

Our study centers on an adapted version of the attention-based spatio-temporal graph convolutional network (ASTGCN) [4], which is designed to capture the complex interplay between traffic flow and network topology via spatial and temporal convolutions enhanced by an integrated attention mechanism. In developing this architecture, we have incorporated specific adaptations to address the unique challenges of urban traffic forecasting as described above.

To provide context for our approach, we compare the adapted ASTGCN against several baseline models. These include graph wavenet (GWNET) [15], hierarchical graph convolutional network (HGCN) [16], graph-refined convolutional network (GRCN) [17], as well as traditional time-series models such as long short-term memory networks (LSTM) [7] and an autoregressive integrated moving average model (ARIMA) [18]. This comprehensive evaluation enables a balanced assessment of the adapted ASTGCN's performance relative to established approaches.

4. Results and Discussion

Our experiments reveal that the volume of training data plays a crucial role in accurately forecasting traffic evolution over the next hour. Using a dataset spanning 27 days collected from 57 camera locations, we evaluated subsets of 3, 7, and 14 consecutive days. Performance was measured using Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root Mean Square Error (RMSE). As shown in Table 1, extending the training period allows the model to better capture the underlying spatio-temporal dynamics. For instance, under the GWNET model, the MAPE decreases from 53.13 % with a 3-day training period to 47.28 % with a 14-day

training period. In the 3-day training scenario, the LSTM model achieves an MAE of 0.0246, a MAPE of 48.18%, and an RMSE of 0.0414, indicating relatively small average errors both in absolute and percentage terms compared to ARIMA, which has higher error values (MAE of 0.0372, MAPE of 81.85 %, and RMSE of 0.0581). Similar comparisons can be drawn across other training durations, where methods like HGCN and GRCN generally show lower errors than ARIMA or even ASTGCN, suggesting they are more effective for this task. The test set consistently includes 7 days, while the validation set comprises 6 days out of the 27-day dataset. Given that traffic dynamics are influenced by phenomena such as phase transitions and hysteresis, precise forecasting of density fluctuations is critical for anticipating transitions between traffic phases, including stop-and-go and synchronized flow [13, 14].

Although our current dataset is limited to 27 days, these findings highlight the potential benefits of scaling up our data volume. Our long-term objective is to compile a full-year dataset, which we expect will yield further insights into traffic evolution and lead to even more robust predictions.

4.1. Inference and Temporal Prediction Accuracy

A further evaluation of our model involves a direct comparison between predicted traffic densities and observed values. Fig. 3 illustrates the temporal evolution of density predictions for the next hour at two randomly selected network nodes. The predictions (depicted in dark red) closely follow the overall trend of the actual density values (depicted in orange), demonstrating the capability of the GNN to capture large-scale temporal patterns.

Closer inspection, however, reveals discrepancies in finer temporal details. This observation, consistent with the findings in [7] – which utilized a larger dataset from a single camera – suggests that the reduced precision at smaller time scales may be attributed to either the limited size of our current dataset or to suboptimal encoding of spatial information within the GNN. Efforts to expand the dataset are underway, and future work will focus on refining the spatial encoding methodology.

Despite common concerns about the computational demands of GNN models, our approach maintains a constant inference time regardless of the training data volume. This is because the model architecture – and hence the number of parameters – remains unchanged as more data is incorporated during training. Consequently, while increased training data enhances prediction accuracy, it does not impose additional computational overhead during inference. Moreover, following [19, 20], our design enables real-time deployment on resource-constrained devices such as a Raspberry Pi, thereby eliminating the need for cloud-based processing.

Table 1. Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root Mean Square Error (RMSE)
evaluated across various model architectures trained on a subset of our data. Data from cameras/intersections
in the Gothenburg network.

Prediction with 3-day training		y training	prediction with 7-day training			prediction with 14-day training			
Method	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE
ARIMA	0.0372	0.8185	0.0581	0.0379	0.8031	0.0557	0.0393	0.8291	0.0572
LSTM	0.0246	0.4818	0.0414	0.0236	0.4689	0.0381	0.0325	0.6024	0.0511
GWNET	0.0259	0.5313	0.0442	0.0247	0.5297	0.0385	0.0245	0.4728	0.0398
HGCN	0.0263	0.4945	0.0414	0.0238	0.4741	0.0380	0.0239	0.4920	0.0376
GRCN	0.0258	0.4720	0.0453	0.0239	0.4927	0.0387	0.0237	0.4815	0.0384
ASTGCN	0.0333	0.6479	0.0516	0.0299	0.5621	0.0466	0.0338	0.5731	0.0543



Fig. 3. Density prediction in next 1 hour (dark red) versus true density in 1 hour (orange) for two different randomly chosen graph locations. Density is measured as space occupied by vehicles within camera mask.

3. Conclusion

In this study we demonstrated that Graph Neural Networks (GNNs) can effectively forecast real-time traffic density using data from Gothenburg's camera network. By leveraging an attention-based spatial-temporal convolutional framework along with hierarchical graph convolutional layers, our model effectively captures both the fine-grained temporal dynamics and the intricate spatial interdependencies present in urban traffic. Vehicle density metrics extracted from real-time images via YOLOv5 - are transformed into structured tensors enriched with temporal encodings, while a dynamically updated, time-dependent Laplacian continuously refines the graph representation to adapt to rapid changes in traffic flow. This integrated approach yields significant improvements in prediction accuracy, as evidenced by the progressive reduction in MAPE when the training set is expanded from 3 to 14 days. Although our current experiments are based on 57 cameras, the inherent design of GNNs makes them well suited to handle much larger networks. In urban environments equipped with thousands of sensors, additional cameras would likely enhance performance by providing richer spatial context, thereby improving

local predictions. At the same time, we recognize that increasing the number of nodes raises computational complexity. This challenge can be addressed through advanced strategies such as efficient graph sampling, intelligent partitioning, and dynamic edge-weighting, through for instance GraphSAGE, which efficiently scale information and selectively focus computational resources on the most informative connections. Evidence from large-scale traffic studies (e.g., those utilizing the PeMSD8 dataset) suggests that the benefits of incorporating a denser sensor network can outweigh the overhead of additional graph complexity.

Acknowledgements

The work of D. L. and A. S. is partially supported by grants from eSSENCE no. 138227, FORMAS no. 2022-151862 and AgTech Sweden. The computations were enabled by resources provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS), partially funded by the Swedish Research Council through grant agreement no. 2022-06725. Thanks to Trafikverket (the transportation authority of Sweden) and the city of Gothenburg for access to their traffic camera data.

References

- [1]. Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, P. S. Yu, A comprehensive survey on graph neural networks, *IEEE Transactions on Neural Networks and Learning Systems*, 2021 Jan, Vol. 32, Issue 1, pp. 4-24.
- [2]. B. Yu, H. Yin, Z. Zhu, Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting, in *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI'18)*, 2018, pp. 3634-3640.
- [3]. W. Jiang, J. Luo, Graph neural network for traffic forecasting: A survey. *Expert Systems with Applications*, Vol. 207, November 2022, 117921.
- [4]. S. Guo, Y. Lin, N. Feng, C. Song, H. Wan, Attention based spatial-temporal graph convolutional networks for traffic flow forecasting, in *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33, Issue 01, Jul. 2019, pp. 922-929.
- [5]. Y. Li, R. Yu, C. Shahabi, Y. Liu, Diffusion convolutional recurrent neural network: Data-driven traffic forecasting, in *Proceedings of the International Conference on Learning Representations (ICLR'18)*, 2018.
- [6]. J. Zhang, Y. Zheng, D. Qi, Deep spatio-temporal residual networks for citywide crowd flows prediction, in *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, Vol. 31, Issue 1, 2017.
- [7]. A. Sopasakis, Traffic demand and longer term forecasting from real-time observations, in *Proceedings of the International Conference on Time Series and Forecasting (ITISE'19)*, 2019, pp. 1247-1259.
- [8]. S. Hochreiter, J. Schmidhuber, Long short-term memory, *Neural Computation*, Vol. 9, Issue 8, 1997, pp. 1735-1780.
- [9]. Ultralytics, Yolov5, https://github.com/ultralytics/ yolov5
- [10]. T.-Y. Lin, M. Maire, S. Belongie, J. Hays, et al., Microsoft COCO: Common objects in context, in Proceedings of the European Conference on Computer Vision (ECCV'14), 2014, pp. 740-755.
- [11]. F. Chung, Spectral Graph Theory, American Mathematical Society, 1997.

[12]. R. Diestel, Graph Theory. Graduate Texts in Mathematics, *Springer*, 2012.

- [13]. A. Sopasakis, Stochastic noise approach to traffic flow modeling, *Physica A: Statistical Mechanics and its Applications*, Vol. 342, Issue 3, 2004, pp. 741-754.
- [14]. A. Sopasakis, M. A. Katsoulakis, Stochastic modeling and simulation of traffic flow: Asymmetric single exclusion process with Arrhenius look-ahead dynamics, *SIAM Journal on Applied Mathematics*, Vol. 66, Issue 3, 2006, pp. 921-944.
- [15]. Z. Wu, S. Pan, G. Long, J. Jiang, C. Zhang, Graph wavenet for deep spatial-temporal graph modeling, in *Proceedings of the 28th International Joint Conference* on Artificial Intelligence (IJCAI'19), 2019, pp. 1907-1913.
- [16]. K. Guo, Y. Hu, Y. Sun, S. Qian, J. Gao, B. Yin. Hierarchical graph convolution network for traffic forecasting, in *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35, Issue 1, May 2021, pp. 151-159.
- [17]. Y. Wei, X. Wang, L. Nie, X. He, T.-S. Chua, Graph-refined convolutional network for multimedia recommendation with implicit feedback, in *Proceedings of the 28th ACM International Conference* on Multimedia, 2020, pp. 3541-3549.
- [18]. B. Williams, L, Hoel, Modeling and forecasting vehicular traffic flow as a seasonal Arima process: Theoretical basis and empirical results, *Journal of Transportation Engineering*, Vol. 129, 2003, pp. 664-672.
- [19]. C. Giannoula, P. Yang, I. Fernandez, J. Yang, et al., Pygim: An efficient graph neural network library for real processing-in-memory architectures, in Proceedings of the ACM on Measurements and Analysis of Computing Systems, Vol. 8, Issue 3, December 2024.
- [20]. F. Vhora, J. Gandhi, A comprehensive survey on mobile edge computing: Challenges, tools. Proceedings applications, in the Fourth of International Conference on Computing Methodologies and Communication (ICCMC'20), 2020, pp. 49-55.

(045)

Knowledge Distillation for Efficient Algerian Dialect Processing: Training Compact BERT Models with DziriBERT

Laggoun Amina ¹, Zakaria Chahnez ¹ and Smaili Kamel ² ¹Higher School of Computer Science (ESI), Algeria ²University Lorraine, LORIA, France Tel.: + 213 783163374 E-mail: ka_laggoun@esi.dz, c_zakaria@esi.dz, smaili@loria.fr

Summary: Dialectal Arabic, particularly Algerian Darija, suffers from a lack of linguistic resources, limiting its integration into natural language processing applications. Additionally, training large language models (LLMs) requires significant computational power. To overcome these challenges, we leverage knowledge distillation to train compact BERT-based student models using DziriBERT as the teacher model. These student models vary in hyperparameter configurations to assess their impact on training effectiveness. As a baseline, we also trained a similarly structured small model without knowledge distillation to measure the technique's contribution. The models were then fine-tuned on multiple downstream tasks, including language and dialect identification, emotion detection, topic classification, and sentiment analysis. The results are promising, with performance comparable to the larger teacher model and even surpassing other multi-dialectal models.

Keywords: Algerian dialect, BERT, Knowledge distillation, Low-resource languages, Small pretrained models.

1. Introduction

Arabic dialects remain an underexplored area of research due to the lack of resources compared to Modern Standard Arabic (MSA), which is widely used in official documents, literature, and education, whereas dialects are primarily spoken in everyday life and used informally on social media. These dialects vary significantly from one country to another and are generally classified into two main groups [1]: Maghrebi dialects, spoken in North Africa and Middle Eastern dialects, which include Levantine dialects, Arabian Peninsula dialects, the Iraqi dialect, as well as Egyptian and Sudanese dialects. Despite their diversity, Arabic dialects lack a standardized representation and writing system, making the development of language models capable of understanding them particularly challenging. With the increasing adoption of Large Language Models (LLMs) such as BERT [2], several efforts have been made to incorporate dialectal Arabic. The first major formal Arabic model, AraBERT [3], was trained on MSA, while later models such as MARBERT [4], Qarib [5], Camel-DA [6], and Camel-Mix [6] integrated a proportion of dialectal Arabic extracted from social media. After that mono dialectal models had been developed specifically for individual dialects, such as DarijaBERT [7] for Moroccan Darija and TunBERT [8] for Tunisian dialect, for the Algerian dialect specifically, the only existing BERT model to date is DziriBERT [9], a BERT based model trained on an Algerian corpus written in both Arabic and Latin scripts. However, a major limitation of these models is their size, as pre-trained language models are computationally expensive, making them difficult to deploy efficiently on resource-limited devices [10]. To address this issue, techniques such as knowledge

distillation have been proposed, allowing a smaller model (student) to learn from a larger model (teacher), thereby benefiting from the teacher's knowledge while maintaining a reduced size. This paper presents our work on applying knowledge distillation to the Algerian dialect, leading to the development of a new lightweight language model called TinyDziriBERT.

The paper is structured as follows: Section 2 provides an overview of related work, Section 3 describes the datasets used for pre-training and evaluation, and Section 4 outlines the methodology. Section 5 details the evaluation process, while Section 6 presents and discusses the results. Finally, Section 7 summarizes the findings of this work.

2. Related Work

Dialectal Arabic is classified as a low-resource language due to the scarcity of labeled datasets necessary for building robust models. Additionally, it lacks a standardized written form and is mainly used in informal settings, such as social media and everyday conversations. In contrast, Modern Standard Arabic (MSA) dominates formal domains like education, literature, and official documents. This imbalance has led to a significant disparity in available resources, making dialectal Arabic an underexplored area of research.

Recent advancements in NLP, particularly the development of large pre-trained language models based on the Transformer architecture [11], have revolutionized the field. Models such as GPT [12], BERT, and RoBERTa [13] have set new benchmarks across various tasks. In the context of Arabic, several models have been proposed to address the challenges posed by the language's diglossic nature. For instance,

AraBERT was among the first major models trained primarily on MSA. Subsequent models, such as MARBERT, Qarib, Camel-DA, and CamelMix, have incorporated substantial amounts of dialectal Arabic data, primarily sourced from social media platforms. Additionally, mono dialectal models like DarijaBERT, MorrBERT, and MorRoBERTa [14] for Moroccan Darija, TunBERT for Tunisian dialect, and DziriBERT for Algerian dialect have been developed to better capture the linguistic nuances of specific dialects.

Despite their effectiveness, these pre-trained models often come with significant computational costs due to their large size, making them impractical for deployment on resource-constrained devices. To mitigate this issue, knowledge distillation has emerged as a promising solution. Originally introduced by Buciluă et al. [15] and later formalized by Hinton et al. [16], knowledge distillation is a model compression technique where a smaller student model is trained to approximate the predictions of a larger teacher model. Since its inception, knowledge distillation has been successfully applied across various domains, including computer vision [17], speech recognition [18], and NLP [19, 20].

In the realm of Arabic NLP, knowledge distillation has been employed to develop lightweight versions of large-scale models. For example, a distillation-based approach has been used for restoring Arabic syntactic diacritics using LSTM networks, where multiple taggers (teachers) were utilized to train a single tagger (student) [21]. Similarly, models like Arabic DistilBERT [22] have been created to reduce BERT model size while maintaining robust performance. Building on these efforts, our work extends the application of knowledge distillation to the Algerian dialect, resulting development in the of TinyDziriBERT – a compact model specifically tailored for this under-resourced variant.

3. Datasets

In the following, we will provide details about the datasets used for training and evaluating our model.

3.1. Training Dataset

The training dataset consists of 1.6 million Algerian dialect sentences extracted from an extended version of CALYOU, an Algerian dialect corpus extracted from YouTube [23], totaling approximately 22 million tokens, written in both Arabic and Latin scripts.

Table 1 presents the distribution of words written in Arabic and Latin scripts within the dataset, while Table 2 presents example sentences from the dataset along with their English translations.

3.2. Evaluation Dataset

The trained models were evaluated using four task specific datasets: Twifil Dataset [24], Narabizi Dataset

[25], Boutef Dataset [26] and another dialectal dataset [27].

Table 1. Word Distribution in the training Dataset.

Arabic Script	Latin Script	Total words
13 M	9 M	22 M

Table 2. Examples of Sentences in the Training Dataset.

Sentence	English Translation
و علاش تبکي يا خويا	Why are you crying my brother
wach had tmskhir	What is this nonsense
mahoumch mndabtin	They are not disciplined

3.2.1. Twifil Dataset

For this dataset, there are two sub-datasets: Twifil Emotion (Twifil-E) and Twifil Sentiment (Twifil-S):

- Twifil Sentiment: 9156 Algerian tweets classified according to the sentiment expressed by the user. The sentiment can be Positive, Negative, or Neutral;
- Twifil Emotion: 5054 Algerian tweets classified into 10 categories based on the Plutchik's model [28].

3.2.2. Narabizi Dataset

The Narabizi Dataset consists of 1287 Algerian sentences written in Latin script, each labeled with both a sentiment and a topic. These sentences were used to create two separate datasets: Narabizi Sentiment (NarabiziS) and Narabizi Topic (NarabiziT). The labels for each dataset are detailed in Table 3.

Table 3. Label Distribution in the Narabizi Datasets.

Narabizi Sentiment	Narabizi Topic
Positive, Negative,	Sport, None, Societal,
Neutral, Mix	Politics, Religion

3.2.3. Boutef Dataset

The Boutef Dataset is a collection of over 3600 fake news posts written in Algerian and Tunisian dialects, Modern Standard Arabic (MSA), French, and English, with occurrences of code-switching between these languages. Each post in the dataset is annotated with 16 tags, including details about the dialect or language used. In our study, we focused on this linguistic information rather than the fake news aspect.

3.2.4. Dialectal Dataset

We used 5271 sentences from this dataset (Dial) [27], categorized as Algerian Arabizi, French, or code switching between the two.

4. Methodology

In this work, we train compact BERT models using knowledge distillation, with DziriBERT as the teacher. Trainings were conducted using the Hugging Face API on a Tesla V100-PCIE-16GB GPU.

While DziriBERT follows the standard BERT-base architecture, the student models, TinyDziriBERT, are significantly smaller. Table 4 provides a detailed comparison of their architectures.

 Table 4. Comparison of Teacher and Student Model

 Architectures.

Model	Number of layers	Hidden size	Attention heads
DziriBERT	12	768	12
TinyDziriBERT	3	256	4

All models share the same vocabulary, consisting of 50000 tokens (the teacher's vocabulary).

Since the higher layers of the teacher model contain richer syntactic, referential, and textual knowledge, the student model is initialized in a bottom-up manner. The teacher has $N \times k$ layers, while the student has Nlayers. The *i*-th layer of the student is initialized using the ($i \times k$)-th layer of the teacher, as shown in Fig. 1.



Fig. 1. Student Model Layer Initialization.

4.1. Training Methodology

The training process utilizes Masked Language Modeling (MLM), where a percentage of tokens in each input sentence from the dataset are randomly masked. The objective is to predict these masked tokens. The sentences with masked tokens are then processed by both the teacher model and the student model to generate logits.

The student model is optimized using a hybrid loss function as shown in Fig. 2, combining the CrossEntropy (CE) loss that measures the difference between the student model's predictions and the true labels and the Kullback-Leibler (KL) Divergence Loss that encourages the student model to mimic the logits generated by the teacher model by comparing their logits.



Fig. 2. Knowledge Distillation for TinyDziriBERT.

The total loss is calculated as:

$$Loss = \alpha \cdot KL(\frac{\hat{y}_{student}}{T} || \frac{\hat{y}_{teacher}}{T}) \times T^{2} + (1 - \alpha) \cdot CE(y_{student}, y_{label}),$$
(1)

where α balances the contribution of the distillation loss and the MLM loss. $\hat{y}_{student}$ and $\hat{y}_{teacher}$ are the logits generated by the student and teacher models, respectively, and y_{label} represents the true tokens. The logits are softened using a temperature parameter *T*. The KL divergence is scaled by T^2 .

4.2. Hyperparameter Variations and Baseline Comparison

To evaluate the impact of various hyperparameters on the model's performance, we conducted a series of experiments by adjusting key parameters. The MLM probability was varied across values of [0.15,...,0.35] to determine the optimal proportion of tokens to mask during training. Additionally, the weighting factor α was tested with values belonging to [0.3, ..., 0.7]. We also experimented with two values for the temperature parameter T, specifically 1 and 2, to analyze the effect of softened logits on the distillation process. In the following, the model name TDBxxyyt refers to TinyDziriBERT, where xx represents the MLM probability, yy denotes the alpha value for distillation, and t indicates the temperature used during knowledge distillation. As a baseline, we trained a TinyDziriBERT model without applying knowledge distillation, using a 50000-token vocabulary created with the WordPiece tokenizer [29] applied to the training dataset.

5. Evaluation

To evaluate performance and efficiency, we fine-tuned the models on Dialect and Language Identification, Sentiment Analysis, Emotion Detection, and Topic Identification. For all the datasets

we split them into 80 % for training and 20 % for testing, with fine-tuning conducted over 4 epochs on an NVIDIA RTX 2060 SUPER (8 GB).

For sentiment analysis, we used the Twifil Sentiment and Narabizi Sentiment datasets. Emotion detection was performed on the Twifil Emotion dataset, while dialect and language identification were conducted using the BOUTEF dataset and the other dialectal dataset (Dial). Topic detection was carried out using the Narabizi Topic dataset.

The models were evaluated against their teacher, DziriBERT, as well as other multi-dialectal models. Table 5 provides a comparison of the sizes of our model with those of other pre-trained models used for benchmarking.

Model	Size (Mo)	Number of parameters (M)	Factor (×)
mBERT	669	167	9
MarBERT	654	163	9
araBERT-v02	543	136	8
Qarib	543	135	8
DziriBERT	498	124.5	7
Camel-Da	439	109	6
Camel-mix	439	109	6
TDB	70.8	18.6	-

Table 5. Models size comparison.

Table 6 presents the vocabulary size of each model along with the script used in the tokens of each vocabulary.

Tabla 6	Vocabular	v size ar	d token	script	ofeach	model
i able o.	vocabular	y size al	ia token	script	of each	model.

Model	Vocabulary size	Token Script
TDB	50K	Arabic + Latin
DziriBERT	50K	Arabic + Latin
mBERT	110K	Multilingual
araBERT-v02	64K	Arabic
MarBERT	100K	Arabic
Qarib	64K	Arabic
Camel-Da/mix	30K	Arabic

6. Results

As presented in Table 7, the distilled TinyDziriBERT model (TDB25052) achieves 78.39 % accuracy on the Twifil Sentiment task, closely approaching DziriBERT's 79.86 %. This highlights the effectiveness of knowledge distillation in retaining much of the teacher model's performance while reducing its size by approximately sevenfold. Likewise, in the Twifil Emotion task, TDB15052 attains an accuracy of 66.04 %. In contrast, the non-distilled TinyDziriBERT (TDB-no distil) exhibits significantly lower performance, with accuracy dropping to 71.08 % for sentiment analysis and 60.01 % for emotion detection. This notable gap underscores

the crucial role of knowledge transfer in preserving model effectiveness.

Table 7. Results on Twifil and Boutef Datas	ets.
---	------

Madal	Twifil		BOU	JTEF	
Model	S-Acc	E-Acc	Acc	F1	
mBERT	73.93	62.47	72.33	70.81	
araBERT-v02	77.39	68.14	77.56	76.27	
Qarib	77.69	70	77.17	76.22	
MarBERT	80.12	70.53	77.92	76.62	
Camel-Da	75.03	67.88	74.77	71.99	
Camel-mix	77.61	69.42	79.30	78.27	
DziriBERT	79.86	70.27	83.32	82.29	
TDB25052	78.39	65.31	78.40	74.84	
TDB15052	77.84	66.04	79.21	75.91	
TDB-nodistil	71.08	60.01	73.05	65.87	

For dialect identification, the best-performing distilled model (TDB35052) achieves an F1 score of 76.12, nearly matching DziriBERT's 77.34 (See Table 8). The non-distilled variant (TDB-no distil) lags behind with an F1 score of 73.19, further validating the role of distillation in enhancing performance.

Table 8. Results on Narabizi and Dial.

Madal	Narabizi		Narabizi		Dial	
widdei	S-Acc	T-Acc	Acc	F1		
TDB25052	56.64	54.10	75.68	75.27		
TDB35052	57.77	53.43	76.32	76.12		
TDB25032	58.16	53.78	75.96	75.78		
TDB-nodistil	50.85	40.15	73.53	73.19		
DziriBERT	64.21	66.36	77.48	77.34		

When compared to larger models like mBERT and AraBERT, TinyDziriBERT demonstrates competitive performance. For instance, mBERT achieves an accuracy of 72.33 % in language identification with Boutef dataset, while AraBERT-v02 reaches 77.56 %, both falling short of TDB15052's 79.21 %. This suggests that models specifically tailored for the Algerian dialect, such as TinyDziriBERT, outperform general-purpose multi-dialectal models in dialect-specific tasks.

The results also reveal that increasing the size of the MLM percentage improves model performance. For example, TDB35052 outperforms TDB25052 with the dialect dataset, achieving an F1 score of 76.12 % compared to 75.27 %.

7. Conclusion

Knowledge distillation proves to be an effective technique for developing compact yet powerful models tailored to the Algerian dialect. By transferring knowledge from a larger, pre-trained model (DziriBERT) to a smaller student model (TinyDziriBERT), we achieve a significant reduction in model size while maintaining high performance across multiple natural language processing (NLP) tasks. Despite its smaller architecture, TinyDziriBERT demonstrates competitive accuracy, closely matching the results of DziriBERT and even surpassing general-purpose models such as mBERT and AraBERT in dialect-specific tasks. TinyDziriBERT opens new possibilities for deploying NLP solutions on mobile devices, edge computing platforms, and low-cost infrastructures for an under-resourced language, namely the Algerian dialect.

References

- S. Harrat, K. Meftouh, M. Abbas, K.-W. Hidouci, K. Smaili, An Algerian dialect: Study and resources, *International Journal of Advanced Computer Science* and Applications, Vol. 7, Issue 3, 2016, pp. 384-396.
- [2]. J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, in *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Vol. 1, June 2019, pp. 4171-4186.
- [3]. W. Antoun, F. Baly, H. Hajj, AraBERT: Transformerbased model for Arabic language understanding, in Proceedings of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools, with a Shared Task on Offensive Language Detection, May 2020, pp. 9-15.
- [4]. M. Abdul-Mageed, A. Elmadany, E. M. B. Nagoudi, ARBERT & MARBERT: Deep bidirectional transformers for Arabic, in *Proceedings of the 59th* Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, Vol. 1, Aug. 2021, pp. 7088-7105.
- [5]. A. Abdelali, S. Hassan, H. Mubarak, K. Darwish, Y. Samih, Pre-training BERT on Arabic tweets: Practical considerations, *arXiv preprint*, 2021, arXiv:2102.10684.
- [6]. G. Inoue, B. Alhafni, N. Baimukan, H. Bouamor, N. Habash, The interplay of variant, size, and task type in Arabic pre-trained language models, in *Proceedings* of the Sixth Arabic Natural Language Processing Workshop, Apr. 2021, pp. 92-104.
- [7]. K. Gaanoun, A. M. Naira, A. Allak, I. Benelallam, DarijaBERT: a step forward in NLP for the written Moroccan dialect, *International Journal of Data Science and Analytics*, Vol. 9, Jan. 2024, pp. 23-40.
- [8]. A. Messaoudi, A. Cheikhrouhou, H. Haddad, N. Ferchichi, M. BenHajhmida, A. Korched, M. Naski, F. Ghriss, A. Kerkeni, Tunbert: Pretrained contextualized text representation for Tunisian dialect, *arXiv preprint*, 2021, arXiv:2111.13138.
- [9]. A. Abdaoui, M. Berrimi, M. Oussalah, A. Moussaoui, Dziribert: a pre-trained language model for the Algerian dialect, arXiv preprint, 2021, arXiv:2109.12346.
- [10]. X. Jiao, Y. Yin, L. Shang, X. Jiang, X. Chen, L. Li, F. Wang, Q. Liu, TinyBERT: Distilling BERT for natural language understanding, in Findings of the Association for Computational Linguistics: EMNLP 2020 (T. Cohn, Y. He, Y. Liu, Eds.), Association for Computational Linguistics, Nov. 2020, pp. 4163-4174.

- [11]. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need, *arXiv preprint*, 2017, arXiv:1706.03762.
- [12]. A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, I. Sutskever, et al., Language models are unsupervised multitask learners, *OpenAI Blog*, Vol. 1, Issue 8, 2019, 9.
- [13]. Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, V. Stoyanov, Roberta: A robustly optimized BERT pretraining approach, *arXiv preprint*, 2019, arXiv:1907.11692.
- [14]. O. Moussaoui, Y. El Younoussi, Pre-training two BERT-like models for Moroccan dialect: MorroBERTA and morrBERT, *MENDEL*, Vol. 29, 2023, pp. 55-61.
- [15]. C. Buciluundefined, R. Caruana, A. Niculescu Mizil, Model compression, in *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'06)*, 2006, pp. 535-541.
- [16]. G. E. Hinton, O. Vinyals, J. Dean, Distilling the knowledge in a neural network, *arXiv preprint*, 2015, arXiv:1503.02531.
- [17]. G. Habib, T. Jan Saleem, S. M. Kaleem, T. Rouf, B. Lall, A comprehensive review of knowledge distillation in computer vision, *arXiv preprint*, 2024, arXiv:2404.00936.
- [18]. J. W. Yoon, H. Lee, H. Y. Kim, W. I. Cho, N. S. Kim, Tutornet: Towards flexible knowledge distillation for end-to-end speech recognition, *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, Vol. 29, 2021, pp. 1626-1638.
- [19]. X. Jiao, Y. Yin, L. Shang, X. Jiang, X. Chen, L. Li, F. Wang, Q. Liu, Tinybert: Distilling BERT for natural language understanding, *arXiv preprint*, 2019, arXiv:1909.10351.
- [20]. V. Sanh, L. Debut, J. Chaumond, T. Wolf, Distilbert, a distilled version of BERT: smaller, faster, cheaper and lighter, *arXiv preprint*, 2019, arXiv:1910.01108.
- [21]. Y. Hifny, Recent advances in Arabic syntactic diacritics restoration, in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'21)*, 2021, pp. 7768-7772.
- [22]. H. Adil, A. Elidrisi, M. Saeed, Knowledge Distillation of BERT Language Model on the Arabic Language, 2023, https://openreview.net/forum?id=bMH1Sk8SSF
- [23]. K. Abidi, M. A. Menacer, K. Smaili, CALYOU: A comparable spoken Algerian corpus harvested from YouTube, in *Proceedings of the 18th Annual Conference of the International Communication Association (Interspeech'17)*, Aug. 2017.
- [24]. L. Moudjari, K. Akli-Astouati, F. Benamara, An Algerian corpus and an annotation platform for opinion and emotion analysis, in *Proceedings of the Twelfth Language Resources and Evaluation Conference*, May 2020, pp. 1202-1210.
- [25]. S. Touileb, J. Barnes, The interplay between language similarity and script on a novel multi-layer Algerian dialect corpus, in Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021 (C. Zong, F. Xia, W. Li, and R. Navigli, Eds.), Aug. 2021, Association for Computational Linguistics, pp. 3700-3712.
- [26]. K. Smaïli, A. Hamza-Jamann, L. David, A. Djegdjiga, BOUTEF: Bolstering Our Understanding Through an

Elaborated Fake News Corpus, in *Proceedings of the* 8th International Conference on Arabic Language Processing, Morocco, Apr. 2024.

- [27]. C. Zakaria, K. Smaïli, B. Sahnoun, A. Chala, R. Agagna, C. Amirat, Algerian Arabizi rumour detection based on morphosyntactic analysis, *International Journal of Knowledge Engineering and Data Mining*, Vol. 8, Issue 1, 2023, pp. 43-66.
- [28]. R. Plutchik, Emotions: A general psychoevoiutionary theory, in Approaches to Emotion, *Psychology Press*, 2014, pp. 197-219.
- [29]. Y. Wu, M. Schuster, Z. Chen, Q. V. Le, et al., Google's neural machine translation system: Bridging the gap between human and machine translation, *arXiv* preprint, 2016, arXiv:1609.08144.

(046)

An Evaluation of General-purpose Large Language Models for Aspect Summarization

S. Frank^{1,2}, C. Gütl¹ and A. Wagner²

 ¹ Graz University of Technology, Institute of Human-Centred Computing, Sandgasse 36, 8010 Graz, Austria
 ² CERN, Esplanade des Particules 1, 1211 Geneva, Switzerland E-mail: sarah.frank@cern.ch

Summary: The rapid rise of large language models (LLMs) has led to the development of both general-purpose and specialized models fine-tuned for concrete tasks. While such specialized models do often lead to some improvements of results, the process of additional training and fine-tuning is resource-intensive, raising concerns about sustainability. This paper takes the first steps in investigating the benefit of specialized versus general-purpose LLMs by focusing on the effectiveness of additional fine-tuning. A small sample of scientific publications and whether the improvements can make up for lack of additional fine-tuning. A small sample of scientific papers was processed using several general-purpose LLMs with different prompts to generate aspect summaries of methods, research questions, and main contributions. The quality of these summaries was assessed with ROUGE scores, with a focus on factual consistency with the original texts. This work provides insights into the successes and limitations of prompt-engineering when opposed to specialized fine-tuned models.

Keywords: Large language models, Aspect summarization, Prompt-engineering, Natural language processing, Scientific summarization.

1. Introduction

The past years have given rise to a rapidly increasing number of large language models (LLMs), both general-purpose as well as fine-tuned, specialized versions. General-purpose models such as GPT, DeepSeek and Llama are intended to cover a broad range of applications and use-cases, whereas fine-tuned models are usually intended to cover a specific type of content, domain, or use-case.

With the training and fine-tuning of models being a laborious process that not only requires extensive training data but is also computationally expensive to an extent that raises sustainability concerns, it is necessary to keep in mind the issue of the computational and energy cost in relation to the benefit of these models. To justify their existence, specialized models that are based on fine-tuning existing general-purpose large language models should add significant value through their use, and recognizably improve results over what can be achieved with general-purpose language models using strategies such as prompt-engineering.

This paper takes a first step into addressing this topic by looking at some of today's most popular general-purpose large language models and evaluating aspect summaries that are created using prompt-engineering to those created by fine-tuned models. For this, a sample of scientific papers contained in the dataset FacetSum [1] is used as input for the language models with a number of different prompts for summaries of used methods, research questions, and main contribution of the paper. The resulting summaries are scored against the reference summaries contained in the datasets using Rouge and compared against the results from the fine-tuned

BART-Facet model [1] and E2E [5] regarding the resulting scores and summary length.

In the following sections, this paper first gives a brief overview of related work in the field of aspect summarization, particularly relating to available datasets and fine-tuned models, followed by a description of the methodology utilized in the course of the research. Subsequently, the results are presented and discussed, with the conclusion wrapping up the paper and proposing some avenues for future work.

2. Related Work

Automatic text summarization is a problem that has led to a multitude of approaches, with the most recent being mostly focused on LLMs. Although general summarization of text is well-researched, aspect summarization – the summarization of text for a specific aspect, e.g. the methods utilized in a scientific article – has been less widely discussed. Both large-scale datasets for this task, specifically in the scientific domain, as well as specifically fine-tuned models, are challenging to source.

In 2021, FacetSum [1] was published as a dataset that had specific tags for "Purpose", "Method", "Findings", and "Value", focusing on the summarization of scientific articles. As the authors stated, large-scale datasets before this tended to be limited in respect to the number of aspects they covered. At the same time, the authors fine-tuned the BART [2] model to summarize text separately for each facet. In the same year, the WikiAsp [3] dataset was published. Although similarly focusing on aspect summaries, this dataset consisted of Wikipedia articles, with aspects created according to common section titles. OASum [4], another dataset based on Wikipedia articles, was published in 2023. With aspects such as "History", "Career", "Background", and "Geography", it likewise sourced them from section titles.

Finally, ACLSum [5], a dataset consisting of scientific papers from the Natural Language Processing (NLP) field, was published in 2024. By utilizing the work of domain experts, the dataset was created without the use of automatically created summaries and specifies the aspects "Challenge", "Approach", and "Outcome". In the same paper, the authors fine-tuned the Llama2 model using full documents as input, with E2E showing promising results.

Beyond these fine-tuned models, aspect summarization is mostly found in other domains, such as summarization of reviews [6] and news articles [7].

3. Methodology

As apparent from the related work, the selection of one or more datasets, as well as fine-tuned models to use for reference presented a challenge in this research. Due to the focus on scientific articles, which typically adhere to a specific structure and require densely concentrated information, datasets drawing from Wikipedia or news articles could not be considered. Both the FacetSum and ACLSum dataset match the requirements, although they consist of different aspects.

For this research, a sample of 30 papers from a variety of domains was taken from the FacetSum dataset for the experiments. At this step, the number of evaluated papers was limited due to the computational cost involved if using the entire dataset. Results from the ACLSum paper were later used in the comparison of Rouge F1-scores. Table 1 shows which considerations were taken to be able to compare results between the different datasets, as well as which wording was chosen to represent the aspect in the query to the language models.

 Table 1. Equivalent aspects between the two datasets

 FacetSum, and ACLSum as well as the term used

 in the query.

Query term	FacetSum	ACLSum
Research purpose	Purpose	Challenge
Research design/ methodology/approach	Method	Approach
Main findings	Findings	Outcome
-	Value	-

For the comparison, some of the most commonly freely available models were used for summarization: GPT-3.5 [8], DeepSeek V3 [9], Mistral's Le Chat [10], as well as Meta- Llama-3.3-70B-Instruct [11]. The Llama model, specifically, was tested using

HuggingChat [12], as well as Poe [13]. For all other models, the dedicated browser versions were used.

The use of the FacetSum dataset for the experiments allowed for the comparison of generated results to the given results to minimize the opportunity for subjective reasoning in the evaluation process. The dedicated paper [1] reported full results regarding Rouge-1, Rouge-2 and Rouge-L scores, which were compared to the results gained through the use of the general-purpose LLMs. In both experiments, as well as ACLSum's E2E approach, the full document text was used as input.

Results were gained through the use of multiple iterations of prompts. Using another small sample of 5 papers, as well as their aspect summaries as target summaries, different wordings for the aspects were tested, as well as small changes in the prompt wording until the content from the reference summaries was represented by the generated summaries. The final term used for each aspect is reported in Table 1.

Occasionally, the LLMs added conversational text or concluding remarks. Any additional text such as "Let me know if you want to know more" or text starting with phrases such as "Overall" or "In conclusion" after the actual aspect summaries was removed and discarded.

4. Results and Discussion

As previously explained, different prompts were selected for trial, varying in results. The creation of the prompts was done manually, with adjustments according to flaws in the results they gave. The final prompt contained a sentence setting the role of the reader as a scientist as this was found to improve results. The final prompt was the following:

We are scientists using a scientific paper for literature research. Summarize the content of this paper specifically and only in regard to "research purpose", "research design/methodology/approach", and "main findings of paper", respectively.

During the calibration period of the prompt, the most common issue was the inclusion of an introduction and conclusion, as well as excessive summary length. Specifically stating the requirement for only the given aspect summaries did not reliably prevent this. As such, this part of the text was discarded as it was not part of the aspect summary sections.

In general, the prompt given above returned promising results for all models and gave clearly delineated summaries for each aspect. However, the average summary length per aspect summary differs significantly by LLM and in comparison, to the reference summary, as well as within each category. Table 2 shows the average number of characters per summary, as well as the standard deviation. While "Purpose" and "Design" summaries are of similar lengths to the reference summaries for most models, "Findings" results in much longer summaries for all but Llama. DeepSeek, particularly, tends to give significantly longer answers in all categories.

results indicate that general-purpose models can reach similar or even better results than the mentioned fine-tuned models with inputs of the full paper text.

	Average summary length in characters (stdev)		
Model	Purpose	Design	Findings
Pafaranca	354.33	341.23	415.13
Kelelelice	(148.61)	(133.64)	(175.87)
CDT	380.80	474.90	770.67
GPT	(76.48)	(166.30)	(225.41)
DeenSeelt	596.30	868.13	1656.43
Беерзеек	(130.01)	(207.85)	(403.81)
Mistral	467.60	541.53	1397.20
Iviisuai	(89.28)	(159.64)	(309.38)
Llama	301.90	404.33	589.70
Liallia	(98.88)	(261.37)	(330.65)

 Table 2. Average summary lengths and their standard deviation per aspect for each LLM.

However, even within the reference summaries, standard deviations are significant. This can be explained by the manual creation of the summaries by the authors themselves, as they are written during the publication process. As such, every author has different expectations and styles for this text, both regarding length and level of detail. Due to the size dataset needed for fine-tuning a language model, this is a natural limitation that is unlikely to be resolved if the target summary is required to be manually created.

As the authors should be the most reliable source regarding the most important points of their paper, this is already the ideal situation, with knowledge of what the authors, themselves, consider the most important takeaways regarding a specific aspect. While this means that it is reasonable to assume that the summaries are factually correct, their length is highly variable. With most language models defaulting to longer answers, this may have an effect on automatic evaluation scores, necessitating human evaluation for results that reliably reflect human opinion.

Results during the experiments showed that particularly for "Findings", more specifications in the prompt may be needed to produce shorter summaries, as the level of detail was excessive in a number of results. For this aspect summary, the standard deviation was also significantly larger for all models except GPT when compared against that of the reference summary.

To further evaluate the generated aspect summaries, Rouge-1, Rouge-2, and Rouge-L F1-scores were calculated. Table 3 shows the averages for each model, as well as the standard deviation in parenthesis. For BART-Facet, the model fine-tuned on the FacetSum dataset [1] and E2E, the Llama model fine-tuned on the ACLSum dataset [5], no standard deviations were given in the respective papers and they are thus missing in the table.

Once again, there was significant standard deviation in the scores. Due to missing reference values from the previous results, it is difficult to draw conclusions and comparisons with them. However, the

Table 3. Average Rouge F1-score	es calculated for the aspect
summaries compared against th	ne reference summaries.

Purpose			
Madal	Rouge-1	Rouge-2	Rouge-L
Widdei	(stdev)	(stdev)	(stdev)
BART-Facet	48.65	27.72	42.55
E2E	30.06	11.33	23.87
CPT	55.04	30.09	42.13
UFI	(11.15)	(14.74)	(14.75)
DeenSeek	51.53	33.78	40.16
Беерзеек	(11.64)	(14.67)	(14.23)
Mistral	53.39	31.69	39.81
Iviisuai	(12.61)	(15.69)	(16.09)
Llama	41.35	14.98	29.95
Liailia	(15.54)	(18.37)	(18.20)
	Desi	ign	
BART-Facet	33.49	11.01	28.07
E2E	44.01	23.03	38.58
CDT	57.16	34.02	45.16
GP1	(15.26)	(18.76)	(18.26)
DeenSeels	41.66	21.13	30.16
Беербеек	(12.91)	(13.05)	(12.76)
Mistral	52.36	30.22	40.44
wiisuai	(13.14)	(13.34)	(13.97)
Llama	33.91	11.79	24.20
Liailia	(13.21)	(14.26)	(12.74)
	Findi	ings	
BART-Facet	34.46	10.49	28.98
E2E	32.85	13.39	27.23
CDT	49.15	27.50	38.08
GP1	(18.19)	(17.51)	(17.48)
DeenSeel	29.09	13.16	19.99
Беербеек	(10.62)	(8.95)	(8.38)
Mistral	34.86	17.48	25.04
Iviisuai	(11.87)	(10.09)	(10.07)
Llama	35.81	11.58	23.04
Liailia	(10.36)	(9.42)	(7.85)

For Bart-Facet, the reported results were generally better when the input was constrained to only the introduction and conclusion (with a Rouge-L score of 43.47/29.07/30.97 for "Purpose", "Method" and "Findings", respectively). Considering this score, BART-Facet outperforms all but GPT's results from this experiment for "Purpose", although all but Llama result in higher scores for "Method" (/"Design"). For "Findings", only GPT outperforms the result. This suggests that it may be beneficial to use only Introduction and Conclusion as input for the summary creation. Doing so would possibly increase the scores shown in Table 3 further, as they did for Bart-Facet [1].

Although GPT consistently performed best for "Design" and "Findings", it was outperformed by Llama when it came to matching the length of the reference summaries, and while Llama performed significantly worse than all other models other than BART-Facet for "Design", it reached average scores for "Findings". E2E, which was based on a previous version of Llama, performed third-highest for "Design".

Finally, both Mistral and DeepSeek returned impressive results for "Purpose" and "Design", but scored significantly lower than GPT's results for "Findings". This coincides with a considerably higher wordcount of Mistral and Deepseek's summaries for "Findings". A specification for maximum summary length could thus particularly influence their scores in this aspect in a positive way.

5. Conclusion and Future Work

This paper evaluated the use of general-purpose LLMs in comparison to two specifically fine-tuned models for the use in aspect summarization for scientific articles. The results showed that although results showed high variance in length and Rouge-scores, some general-purpose language models can reach comparable results with well-tailored prompts.

This result indicates that comparisons between fine-tuned models and their base-models must be accompanied with comparisons to popular general-purpose LLMs to show that the increased quality justifies the additional time and energy that is spent on fine-tuning models. It suggests that prompt engineering has a significant influence on result quality that can potentially be used to make up for the lack of task-specificity of a model.

However, while the results presented in this paper are promising, certain limitations are present. One of the referenced fine-tuned models was published in 2021 [1] and as such a fine-tuned version of a more recent base model may produce improved results. Similarly, E2E [5] is based on Llama 2, a model that has since likewise been replaced by more recent versions. Furthermore, although this paper supports the critical evaluation of fine-tuned models versus the use of prompt engineering, future work is needed to extend the experiments with a larger dataset for the comparison of the summaries. Similarly, further models should be considered such as Claude and Gemini.

Finally, the comparison of a fine-tuned model with the base model it is based on, once again taking care to

optimize the prompt for the task, may allow for further conclusions regarding prompt-engineering's impact on the quality of results.

References

- R. Meng, et al., Bringing structure into summaries: a faceted summarization dataset for long scientific documents, in *Proceedings of the 59th Annual Meeting* of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, Vol. 2, 2021, pp. 1080-1089.
- [2]. M. Lewis, et al., BART: denoising sequence-tosequence pre-training for natural language generation, translation, and comprehension, in *Proceedings of the* 58th Annual Meeting of the Association for Computational Linguistics, 2020, pp. 7871-7880.
- [3]. H. Hayashi, et al., WikiAsp: A dataset for multi-domain aspect-based summarization, *Transactions of the Association for Computational Linguistics*, Vol. 9, 2021, pp. 211-225.
- [4]. X. Yang, et al., OASum: large-scale open domain aspect-based summarization. in Findings of the Association for Computational Linguistics: ACL 2023, *Association for Computational Linguistics*, 2023, pp. 4381-4401.
- [5]. S. Takeshita, T. Green, I. Reinig, K. Eckert, S. Ponzetto, ACLSum: a new dataset for aspect-based summarization of scientific publications, in Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Vol. 1, 2024, pp. 6660–6675.
- [6]. L. S. d. Souza, M. G. Manzato, Aspect-based summarization: an approach with different levels of details to explain recommendations, in *Proceedings of the Brazilian Symposium on Multimedia and the Web* (WebMedia'22), 2022, pp. 202-210.
- [7]. B. Tan, L. Qin, E. Xing, Z. Hu, Summarizing text on any aspects: a knowledge-informed weakly-supervised approach, in *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (*EMNLP'20*), 2020, pp. 6301-6309.
- [8]. OpenAI, GPT-3.5, https://openai.com/
- [9]. A. Liu., et al., Deepseek-v3 technical report, *arXiv* preprint, 2024, arXiv:2412.19437.
- [10]. Mistral, https://mistral.ai/
- [11]. A. Dubey, et al., The llama 3 herd of models, *arXiv* preprint, 2024, arXiv:2407.21783.
- [12]. HuggingChat, https://huggingface.co/chat/
- [13]. Poe, https://poe.com

(047)

Characteristics of Dynamic Velocity Response in Hand Movements Using Frequency and Time Modeling Techniques

<u>C. L. Sandoval-Rodriguez</u>^{1,2}, A. F. Jimenez-Quezada¹, N. Orejarena-Osorio¹, O. Lengerke¹, and D. M. Reyes-Bravo³

¹ Unidades Tecnológicas de Santander, Calle de los estudiantes, 9-82, 680005, Bucaramanga, Colombia ² University of the Basque Country UPV/EHU, Plaza Torres Quevedo 1, 48013 Bilbao, Spain ³ Universidad Autonoma de Bucaramanga, Av. 42 #48 – 11, 680002, Bucaramanga, Colombia Tel.: (+57) 607 6917700 E-mail: csandoval@correo.uts.edu.co

Summary: Natural human hand movements are a topic of current research. Tools have been used to analyze surface electromyographic signals -sEMG- associated to each movement and their main characteristics. The motivation of this work is to show a systematic study of the possible mathematical relationships between the velocity exerted in basic human hand movements and the electrical activity of the muscles that produce the movements, based on surface electromyography. This study evaluates 14 healthy subjects and six types of movements (pronation, supination, ulnar deviation, radial deviation, flexion and extension -84 SEMG and velocity recordings). The work is divided into 2 parts. First, a model in the Laplace domain is proposed that relates these two variables (velocity and sEMG) using the sEMG signal envelope. Secondly, a linear model (time domain) focused on predicting the velocity of each movement is obtained. The results show strongly linear models in the time domain, with high differences between each type of motion, high coefficients of determination (0.95 on average) and an MSE of 21.18 %. In the Laplace analysis, second order models predominate with best fit characteristics of 82.45 % on average for all analyzed motions. The velocity response characteristics show relatively low response times, but high steady-state setpoint tracking errors.

Keywords: Human hand motions, Linear models, Dynamic response, Velocity of movements, sEMG.

1. Introduction

Many previous works have sought to relate the surface electromyography (sEMG) signal to the characteristics (velocity and force) of the kinematics of human hand motion [1-3]. Others have considered the problem only as a classification issue with good results [4-6]. Techniques for feature extraction from surface electromyography signals have been employed to determine the type of motion, both in the time domain and in the frequency domain [7-13]. However, the research is still under development. Looking for human hand movements to approximate natural movements, scientists have employed different alternatives that allow relating force and velocity to the electrical activity of the muscles [14-18]. However, the great majority of works are related to the classification of the type of movement and not to its attributes. The contribution of this systematic work is directed at relating the velocity exerted in six human hand movements (pronation, supination, flexion, extension, ulnar deviation and radial deviation) and the sEMG signals obtained in the proximal third of the forearm in 14 healthy subjects. The paper is divided into three main parts [19-20].

The methodology used explains how the information was collected, the treatment performed to the velocity and electromyographic signals to obtain their envelope. Also, the modeling performed in the frequency domain where basic first, second and third order process models were used. Additionally, the modeling performed in the time domain. Then in the results section you can see the model fitting error calculated as Mean Squared Error (MSE) for the models in frequency in each order. It will also be possible to observe the performance evaluation of the models calculated in open loop to estimate response times and steady state errors from the calculated models. Another relevant aspect that will be evidenced in the results is the possible linear relationship in the time domain reporting an overall MSE of 21.18 %. Finally, the section on conclusions derived from the work.

2. Materials and Methods

2.1. Experimentation and Data Collection

The database has 84 velocity and sEMG recordings taken from 14 healthy subjects. We calculated the sEMG envelope using moving average envelopment. The following movements were studied: flexion-extension, supination pronation, ulnar deviation-radial deviation.

Research ethics approval was obtained by the ethical approval to report this case obtained from * *Ethics Committee for Research, Bioethics and Scientific Integrity – CEI Resolución 02-474 de agosto 4 del año 2021/ FIN 11-15.*

The process shown in Fig. 1 was used, using a sampling rate of 960 frames/s to calculate the velocity profile.



Fig. 1. Sequence to obtain the velocity profiles.

We obtained the sEMG envelope using a 2-sample moving average over 20 ms windows. Fig. 2 shows an example of the similarity of both signals (sEMG envelope and velocity profile).



Fig. 2. Top panel sEMG envelope, bottom panel velocity profile in the flexion-extension motion.

2.2. Complex Frequency Domain Modeling

The models were obtained using the parametric system identification technique [17].

A general approximation model was proposed as in equation (1).

$$G(s) = k \frac{(1+Tz*s)e^{-Td*s}}{(1+Tp1*s)(1+Tp2*s)(1+Tp3*s)}, \qquad (1)$$

where K is a Gain, Tz is a Time from zero, Td is a Delay time, Tp1 is the Pole 1 time, Tp2 is the Pole 2 time, Tp3 is the Pole 3 time.

Note: The predominant model for the present study contemplates a second-order structure.

2.3. Statistical Analysis

In order to obtain a general model, the model parameters were averaged according to equation (2).

$$x' = \frac{\sum_{n=1}^{N} x(n)}{N}$$
(2)

Dispersion was analyzed by calculating the standard deviation according to equation (3).

$$\sigma = \sqrt{\frac{\sum_{n=1}^{N} (x(n) - x')^2}{N-1}},$$
(3)

and the coefficient of variation (CV). According to equation (4).

$$CV = \frac{x}{\sigma} * 100\% \tag{4}$$

2.4. Estimation of the Model Performance

For performance estimation, the settling time, as an estimator of the model response speed, and the steady-state error at a step input were calculated to evaluate the tracking capability to the setpoint. For the settling time, the 2 % criterion was used [17, 18].

2.5. Time Domain Modeling

We proposed a linear relationship between the velocity and the sEMG tone signal (envelope) using the proposed model according to equation (5), where a and b are model parameters and t is the time variable. In Fig. 3, you can see the possible relationship between the velocity profile and the envelope sEMG signals.

$$Velocidad(t) = a * Tono_{sEmg(t)} + b$$
 (5)

We report results as interquartile ranges of a and b. We calculated the goodness of fit of the model from the coefficient of determination R2, the results in Table 5. Curve fitting was applied for each subject and for each movement.

3. Results

The results are divided into 3 parts. First the modeling results in the Laplace domain. Second, the model performance results (response speed and tracking capability via steady state error). Finally, the results of the linear model in the time domain.

3.1. Complex Frequency Domain Modeling Results

As explained above, the general model proposed is the one corresponding to equation 1. Tables 1, 2 and 3 show the results for each order (1, 2 and 3) respectively.

From Tables 1-3 it can be seen that the best performance with respect to the percentage of fit, standard deviation and average MSE in all movements is obtained for order 2.

3.2. Performance Analysis of the Obtained Model

Taking into account that the best fitting behavior of the model was obtained for a 2-pole system (second order). For the analysis, second order models were taken for each movement and settling times and steady state error were estimated for a unitary step input. This hypothesis seeks to estimate the performance that the system composed of the set of soft tissues together with the bone structure would have to respond to a brain order of abrupt change in the speed of each movement. An example for the pronation movement can be seen in Fig. 3.

Movement	% error	Standard deviation
Pronation	21.83	9.3
Supination	19.04	7.5
Flexion	19.95	9.4
Extension	22.87	11.2
Radial Deviation	18.32	4.3
Ulnar Deviation	17.79	7.1
Average	MSE 19.95 %	8.13

Table 1. Comparison of the models obtained for first orde	er
in the different movements analyzed.	

Table 2. Comparison of the models obtained for secon	١d
order in the different movements analyzed.	

Movement	adjustment	Standard deviation
Pronation	21.37	6.4
Supination	17.35	6.8
Flexion	16.17	5.4
Extension	22.47	8.8
Radial Deviation	18.14	4.6
Ulnar Deviation	15.79	9.8
Average	MSE 18.55 %	6.97

Table 3. Comparison of the models obtained for third order in the different movements analyzed.

Movement	% adjustment	Standard deviation
Pronation	29.18	27.1
Supination	24.93	5.7
Flexion	14.32	8.4
Extension	21.97	12.7
Radial Deviation	18.88	7.7
Ulnar Deviation	19.77	17.9
Average	MSE 21.51 %	13.25

The summary results are shown in Table 4.

Table 4. Performance estimation resultsfor the second-order model.

Movement	Orden 2 (ts, ESS) [s,%]
Pronation	[1.5,47]
Supination	[2.1,72]
Flexion	[1.3,14]
Extension	[2.3,27]
Radial Deviation	[4.15,22]
Ulnar Deviation	[6.4,38]
Average	[2.96, 36.5]

A good performance is observed in terms of response speed, as a function of settling time, however, the average error of 36.5 % shows the prevailing need for adaptation to correct this error. That is to say, the calculated models have an open-loop behavior with high setpoint tracking errors, this would be an impediment for the development of an arm that replicates natural movements. A relationship in the time domain is then sought in order to mitigate this in the next section.



Fig. 3. Response to a unitary step-type input, for a healthy subject, in the pronation movement. A settling time of 0.826 (s) and a final value of 0.0376 are observed, showing a steady state error of approximately 96.24 %.

3.3. Time-domain Modeling Results

Taking into account the results obtained in the complex frequency domain, a linear temporal relationship as in equation (5) is sought. A scatter plot relating the sEMG signal envelope to the velocity profile for a healthy patient in the flexion movement can be seen in Fig. 4.



Fig. 4. Linear relationship between the velocity of the bending movement and the sEMG envelope for the same movement, for a healthy patient.

Table 5 shows the results as interquartile range and median of the model parameters of equation (5) and the coefficient of determination.

Movement	a median (IQR)	b median (IQR)	R ²
Pronation	4536 (4448, 4624)	23.93 (19.54, 28.31)	0.9092
Supination	0.451 (0.445, 0.458)	106.3 (103.9, 108.7)	0.9468
Flexion	10.18 (10.08, 10.29)	-51.1 (-53.9, -48.3)	0.9714
Extension	1056 (1044, 1069)	-118.1 (-122.4, -113.9)	0.9648
Radial Deviation	3.32 (3.28, 3.36)	49.88 (48.87, 50.88)	0.9711
Ulnar Deviation	5.47 (5.39, 5.56)	-70.1 (-72.8, -67.5)	0.9369

 Table 5. Summary of results of the linear regression models for each type of movement.

We used the MSE as a metric to estimate the model error in the validation process. Fig. 5 shows an example (pronation) of the output model behavior and compares it to the original measured tone. We obtained MSE globally for all samples (84 records and signals from the output model). The MSE was 21.18 %.



Fig. 5. Comparison of model output (red color) and actual pronation movement velocity for a healthy subject.

4. Discussion

There are methods that could reduce the error of the response speed of the second-order model. One of them is the use of adaptive filters, which allow the system parameters to be dynamically adjusted according to variations in the sEMG signals, thus improving the accuracy of the model. In addition, the incorporation of feedback loops can be crucial, as these systems continuously compare the model output with the desired reference, allowing real-time corrections and significantly reducing error. These strategies not only optimize the stability and response speed of the system, but also increase its robustness to external disturbances and variations in model parameters, achieving greater accuracy and efficiency in hand motion control based on sEMG signals [21].

On the other hand, due to the variability between subjects, the creation of a universal model is not simple, nevertheless, for practical applications it could be calibrated for each subject. Additionally, a universal model could be generated with the use of machine learning and specific models for different subjects [22].

Another valuable prospect for future research could be deep learning approaches, such as recurrent neural networks (LSTMs) and convolutional neural networks (CNNs), which are able to learn more complex representations and capture temporal and spatial patterns in sEMG signals, which may result in increased accuracy and robustness in hand velocity prediction. The integration of these advanced approaches could significantly improve the accuracy and efficiency of hand motion control [21], providing a solid foundation for future research and applications in rehabilitation and human-machine interfaces.

Finally, the robustness of the models in the time domain can be assessed through the Cross-validation technique [23], which allows to observe if the error is consistent and if the model is reliable.

5. Conclusions

From the results obtained with the model in the complex frequency domain, it is observed that the best behavior occurs for systems of order 2. When analyzing the performance of the model obtained, it is observed that the steady state errors, in the presence of an abrupt change input, are very high on average. This indicates that it requires adaptation, which could be corrected in future research by adapting a control action to reduce this. Nevertheless, the response times are below 3 seconds.

On the other hand, the coefficients of variation when assessing the fit of the 14 subjects are high. Therefore, the idea of generalization of the model is still low. Thus, the methodology applied here should be performed for each subject separately. Also, the models calculated by this strategy present high errors and adaptation times that, although low, in the context of a motion control can be high, so that a control strategy should be integrated to mitigate these deviations.

Finally, the relationship between surface electromyography tone and velocity profiles in the time domain is strongly linear. However, there is a scattered behavior of the linear model parameters. In addition, the difference between subjects and type of movement generates different parameters in each model with considerable scatter. Therefore, the methodology must be used in each situation, which makes it difficult to apply a general model in hardware that replicates the velocities estimated from the sEMG treatment. However, the overall MSE has not been high and could be compensated with a control system that rejects these uncertainties.

References

[1]. K. Englehart, B. Hudgins, P. A. Parker, M. Stevenson, Classification of the myoelectric signal using time-frequency based representations, *Medical Engineering & Physics*, Vol. 21, Issue 6-7, 1999, pp. 431-438.

- [2]. L. F. Bender, Muscles alive: their functions revealed by electromyography, JAMA, Vol. 201, Issue 4, 1967, 277.
- [3]. R. Álvarez Fiallo, C. Santos Anzorandia, E. Medina Herrera, Diagnóstico electromiográfico de las enfermedades neuromusculares, *Revista Cubana de Medicina Militar*, Vol. 36, Issue 1, 2007.
- [4]. L. Gila Useros, A. Malanda Trigueros, I. Rodriguez Carreño, J. Rodriguez Falces, J. Navallas Irujo, Métodos de procesamiento y análisis de señales electromiográficas, *Anales del Sistema Sanitario de Navarra*, Vol. 32 (Supl. 3), 2009, pp. 27-43.
- [5]. A. Waris, I. K. Niazi, M. Jamil, K. Englehart, W. Jensen, E. N. Kamavuako, Multiday evaluation of techniques for EMG-based classification of hand motions, *IEEE Journal of Biomedical and Health Informatics*, Vol. 23, Issue 4, 2018, pp. 1526-1534.
- [6]. Y. Sun, et al., Intelligent human computer interaction based on non redundant EMG signal, *Alexandria Engineering Journal*, Vol. 59, Issue 3, 2020, pp. 1149-1157.
- [7]. S. Abbaspour, A. Naber, M. Ortiz-Catalan, H. GholamHosseini, M. Lindén, Real-time and offline evaluation of myoelectric pattern recognition for the decoding of hand movements, *Sensors*, Vol. 21, Issue 16, 2021, 5677.
- [8]. B. Kundu, D. S. Naidu, Classification and feature extraction of different hand movements from EMG signal using machine leaning based algorithms, in *Proceedings of the International Conference on Electrical, Communication, and Computer Engineering (ICECCE'21)*, 2021, pp. 1-5.
- [9]. M. B. I. Reaz, M. S. Hussain, F. Mohd-Yasin, Techniques of EMG signal analysis: detection, processing, classification and applications, *Biological Procedures Online*, Vol. 8, 2006, pp. 11-35.
- [10]. X. Ren, Y. G. Soo, M. Odagaki, F. Duan, sEMG-based hand motion recognition system using RMSR and AR model, in *Proceedings of the 36th Chinese Control Conference (CCC'17)*, 2017, pp. 5410-5415.
- [11]. C. M. Durán Acevedo, A. L. Jaimes Mogollón, Optimización y clasificación de señales EMG a través de métodos de reconocimiento de patrones, *Iteckne*, Vol. 10, Issue 1, 2013, pp. 67-76.
- [12]. W. Seok, Y. Kim, C. Park, Pattern recognition of human arm movement using deep reinforcement learning, in *Proceedings of the International Conference on Information Networking (ICOIN'18)*, 2018, pp. 917-919.
- [13]. M. Kim, K. Kim, W. K. Chung, Simple and fast compensation of sEMG interface rotation for robust

hand motion recognition, *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, Vol. 26, Issue 12, 2018, pp. 2397-2406.

- [14]. L. A. Hidalgo Torres, Y. San Martin Reyes, J. D. Chailloux Peguero, Capture of the Voluntary Motor Intention from the Electromyography Signal, in Proceedings of the VIII Latin American Conference on Biomedical Engineering and XLII National Conference on Biomedical Engineering (CLAIB-CNIB'19), October 2-5, 2019, Cancún, México, pp. 28-36.
- [15]. E. Guzmán-Muñoz, G. Méndez-Rebolledo, Electromyography in the rehabilitation sciences, *Revista Salud Uninorte*, Vol. 34, Issue 3, 2018, pp. 753-765.
- [16]. S. Nebe, et al., Enhancing precision in human neuroscience, *Neuroscience*, 2023.
- [17]. C. L. Sandoval-Rodriguez, A. C. Pita-Mejia, R. Villamizar-Mejia, B. E. Tarazona-Romero, O. Lengerke-Perez, Model to relationship the speed of hand movements with the SEMG signals from the forearm, *Journal of Physics: Conference Series*, Vol. 2224, 2022, 012094.
- [18]. C. L. S. Rodriguez, R. V. Mejia, B. E. T. Romero, A. D. R. Quintero, A. J. R. Nieves, Relationship between force signal and superficial electromyographic signals associated to hand movements, *Periodicals of Engineering and Natural Sciences*, Vol. 11, Issue 1, 2023, pp. 64-73.
- [19]. A. D. Rincon-Quintero, C. L. Sandoval-Rodríguez, N. A. Castillo-Zambrano, A. F. Jiménez-Quezada, O. Lengerke-Perez, C. A. Angulo-Julio, Trends in associated techniques for Pattern Recognition for the classification of hand movements with electromyography signals, in *Proceedings of the Anais* do XV Simpósio de Engenharia Biomédica, 2023.
- [20]. C. L. Sandoval-Rodríguez, C. J. Arizmendi-Pereira, D. M. Reyes-Bravo, O. Lengerke, R. Palacio, A. F. Jimenez-Quezada, Differences in the force exerted during wrist movements are explained by a general mathematical model, *Heritage and Sustainable Development*, Vol. 6, Issue 2, 2024, pp. 749-756.
- [21]. A. Waris, I. K. Niazi, M. Jamil, K. Englehart, W. Jensen, E. N. Kamavuako, multiday evaluation of techniques for EMG-Based classification of hand motions, *IEEE J. Biomed. Health Inform.*, Vol. 23, Issue 4, Jul. 2019, pp. 1526-1534.
- [22]. J. M. Pineda, Modelos predictivos en salud basados en aprendizaje de maquina (machine learning), *Revista Médica Clínica Las Condes*, Vol. 33, Issue 6, Nov. 2022, pp. 583-590.
- [23]. J. Allgaier, R. Pryss, Cross-Validation Visualized: A narrative guide to advanced methods, *Mach. Learn. Knowl. Extr.*, Vol. 6, Issue 2, Jun. 2024, pp. 1378-1388.

(048)

Graphical User Interface for Volumetric Capnography: Parameter **Estimation and Fowler's Method Implementation**

C. L. Sandoval-Rodriguez^{1,2}, N. Orejarena-Osorio¹, A. F. Jimenez-Quezada¹, and O. Lengerke¹

¹Unidades Tecnológicas de Santander, Calle de los estudiantes 9-82, 680005, Bucaramanga, Colombia ² University of the Basque Country UPV/EHU, Plaza Torres Quevedo 1, 48013 Bilbao, Spain Tel.: (+57) 607 6917700

E-mail: csandoval@correo.uts.edu.co

Summary: This study presents the development of a Graphical User Interface (GUI) using MATLAB, incorporating Curve Fitting and GUIDE tools to analyze volumetric capnography (VC) data. The primary objective is to estimate essential respiratory parameters, including anatomical dead volume (VDaw) and end-expiratory CO₂ volume (ETCO₂), by implementing Fowler's method for volumetric capnography analysis. The proposed GUI automates the graphing process and ensures accurate parameter identification from a patient database, offering a valuable resource for clinical applications. Volumetric capnography is a non-invasive technique used to monitor ventilation efficiency and pulmonary function. In this study, the developed software was applied to a dataset consisting of previous patient measurements, effectively generating automatic capnograms and extracting key physiological variables. The methodology involved signal preprocessing, curve fitting using the Levenberg-Marquardt algorithm, and parameter estimation via numerical integration techniques. Results demonstrated the GUI's effectiveness in delivering reliable VDaw and ETCO₂ measurements. The interface's user-friendly design allows clinicians to intuitively analyze capnograms and monitor respiratory status in real time. This application has significant potential in enhancing decision-making during mechanical ventilation management and cardiopulmonary resuscitation scenarios. Future work will focus on refining the algorithm's robustness, reducing noise interference, and expanding the tool's capabilities for real-time monitoring in intensive care units. The developed software represents a step towards advanced respiratory diagnostics and improved patient outcomes.

Keywords: Volumetric capnography, CO₂ volume, Cardiopulmonary resuscitation, Fowler's method, Exhalation.

1. Introduction

The pandemic caused by the coronavirus generates in people acute respiratory infection causing respiratory distress being one of the most common complications and if allowed to progress can produce respiratory arrest in patients [1]; from the World Health Organization (WHO) emphasis is placed on improving the processes of mechanical ventilation and CPR maneuvers in order to mitigate human losses [2-5].

CPR maneuvers are used to improve patient prognosis and are widely used in critically ill patients during the pandemic [6, 7], is performed in cases where the endotracheal tube (ETT) is not in the correct position, making it impossible for air to enter the lungs or even in the correct position of the ETT, sometimes the frequency and volume of ventilations is not the required one; likewise, assisted ventilation is not applied with quality so that the patient can be stabilized. [8-10], therefore, these variables must be monitored and other variables must be added to guarantee the quality of ventilation and compressions to the patient [11].

Currently, volumetric capnography provides health professionals with tools that monitor information on the distribution of air entering the airways, helps to calculate the volume of carbon dioxide in a tidal volume [12], being very useful to assess the relationship between ventilation and perfusion, allowing the detection of different states or clinical activities such as lung collapse [13], acute respiratory distress syndrome, detection of re-inhalation, metabolic activity in patients with assisted ventilation or the return of spontaneous circulation, an important factor during the application of CPR maneuvers, and thus determine the effectiveness and quality of ventilation [14-17].

The methods used relate the theoretical concepts of volumetric capnography, Fowler's method for capnography analysis, mechanical ventilation, cardiopulmonary resuscitation maneuvers, MATLAB software programming language, curve fitting and numerical integration techniques.

The development of this Graphical User Interface [18] in MATLAB software will allow to obtain volumetric capnographies automatically, which will provide preliminary information of the patients according to their shape and values. It will also apply Fowler's method to each exhalation individually extracting information regarding the anatomical dead volume and the CO volume at the end of the exhalation, presenting an average of these values at the end of the analysis of the whole signal [19].

2. Signal Reception with Less Interference

The complete signal obtained from the database is plotted, generating on its right side a signal with distortion due to the chest compressions applied to the patient (see Fig. 1), the cleanest section observed in Fig. 2 is chosen as a reference.



Fig. 2. Interference-free signal.

3. Curve Fitting

To perform the curve fitting, the data that compose the expirations are separated individually to be introduced in the Curve Fitting tool of MATLAB, which generates the individual expiratory curve shown in Fig. 3.



Fig. 3. Individual exhalation

The generated signal contains noise that does not allow to obtain a clean signal; to improve the signal there are several methods to calculate the anatomical dead space from the volumetric capnography [20], when this mathematical model is obtained at a general level, it is taken to the Curve Fitting tool with the signal data, where the coefficients of the model that best fit the experimental data are found (see Fig. 4).



Fig. 4. Curve fitting result.

4. Areas P and Q

One of the methods for the analysis of volumetric capnographies and the identification of dead anatomical volume is the Fowler method, which consists of equating the areas of volumetric capnographies P and Q.

The area Q is made by means of the function and its respective coefficients that describe the signal, applying an integration defined by the lower limits will be the beginning of the expiration and the upper limit will be the variation to equalize the areas; for the area P is calculated the result of the linear regression of data that compose the middle zone of phase III of the expiration and the integral of the model of the signal function.



Fig. 5. Areas P y Q.

5. Results

By implementing the curve-fitting method, based on the Levenberg-Marquardt algorithm [21, 22], parameterized by the chosen VC mathematical model, to the experimental data of the five patients, it was observed that the coefficients of the characteristic equation reflected favorable goodness-of-fit values, according to the data of "Percent of confidence bounds", "Adjusted R-square" and root mean square error (RMSE), given by the MATLAB software.

The summary of the goodness of fit values, obtained through the curve fitting tool from MATLAB, for each patient, is shown in Table 1.

Patient/ Episode	% Confidence	Adjusted R ²	RMSE
1	95 %	0.997	1.23
2	95 %	0.9994	0.6026
3	95 %	0.9995	0.6067
4	95 %	0.9964	0.2551
5	95 %	0.9943	0.2437

Table 1. Results curve fitting.



Fig. 6. Curve fitting patient 1.



Fig. 7. Curve fitting results patient 1.



Fig. 8. Calculation and averaging of the values of the selected exhalation.

At the end of the development of the interface, the correct functioning of all the actions included in the program applied to the signals of all the patients and their respective exhalations was verified.

The average data of anatomical dead volume and end-expiratory volume obtained from analyzing each of the six expirations of each patient, using this computational tool, are shown in Table 2.

Table 2. Average Data Obtained.

Patient	Average VDaw (ml)	Average ETCO2 (mmHg)
1	133.41	42.623
2	129.56	52.425
3	128.67	56.426
4	127.65	9.6
5	146.84	7.4

6. Conclusions

The structure in which the software was designed and the Graphical User Interface in MATLAB allows a dynamic use of the program, providing an orderly access to the information of each patient, achieving the analysis of each signal very intuitively. The general model of the function used in the curve fitting algorithm does not fit well in some signals with peak values of CO Volume in phase III of expiration, this impacts the accuracy of the data. For continue advancing in the development of tools that allow the study of volumetric capnographies and to make the leap from off-line analysis to on-line analysis, it is necessary to deepen in the design and application of filters that eliminate interferences in the input signal.

7. Limitations

This interface performs signal analysis offline. To enhance its clinical utility and enable real-time application, future versions could incorporate modules for real-time data acquisition and processing from ventilators or capnometers. Such functionality would allow healthcare professionals to make faster and more informed decisions during procedures like CPR.

A notable limitation of this study is the use of only five patient datasets. However, it is important to emphasize that this work serves as a pilot test to assess the feasibility of the proposed interface. Additionally, acquiring datasets with concurrent capnography and volume information is challenging due to the limited availability of such records. Future research should consider including a larger and more diverse patient sample to strengthen the validity of the results.

Furthermore, this study did not conduct a direct comparison with commercial capnography software. This decision was primarily influenced by the proprietary nature and high cost of existing commercial solutions. Future studies could address this gap by performing comparative evaluations with commercial tools to validate the system's performance and assess its potential for clinical implementation.

Another key challenge in interpreting capnography signals during CPR is the presence of artifacts caused by chest compressions. Previous studies have shown that artifacts are more prominent during manual compressions, whereas mechanical systems produce significantly fewer disturbances [11]. To mitigate this issue, future implementations could integrate advanced filtering techniques, such as wavelet-based noise removal or adaptive filtering.

Finally, while the proposed curve-fitting approach demonstrated good performance in the conducted tests, its generalizability to larger patient populations may be limited. To reduce the risk of overfitting, it is advisable to explore alternative nonlinear models or apply crossvalidation techniques using additional datasets. This would provide a more comprehensive evaluation of the model's robustness across different clinical scenarios.

References

- [1]. C. Calvo, M. G. López-Hortelano, J. C. de Carlos Vicente, J. L. V. Martinez, G. de trabajo de la Asociación, et al., Recommendations on the clinical management of the COVID-19 infection by the new coronavirus SARS-CoV2. Spanish Paediatric Association working group, *An. Pediatra (English Ed.)*, Vol. 92, Issue 4, 2020.
- [2]. O. McAlister, et al., cpr guideline chest compression depths may exceed requirements for optimal physiological response, in Proceedings of the Computing in Cardiology Conference (CinC'18), 2018, Vol. 45, pp. 1-4.
- [3]. N. Segal, *et al.*, Correlation of end tidal carbon dioxide, amplitude spectrum area, and coronary perfusion pressure in a porcine model of cardiac arrest, *Physiol. Rep.*, Vol. 5, Issue 17, 2017, e13401.
- [4]. S. Blanco Lorenzo, Capnografia en las RCPs: conocimiento y valoración de este sistema de monitorización por profesionales de enfermeria, http://hdl.handle.net/2183/26145
- [5]. A. R. Panchal, *et al.*, Part 3: adult basic and advanced life support: 2020 American Heart Association guidelines for cardiopulmonary resuscitation and emergency cardiovascular care, *Circulation*, Vol. 142, Issue 16, 2020,, pp. S366-S468.
- [6]. M. Aliaño Piña, C. Ruiz Villén, J. Galán Serrano, P. Monedero Rodríguez, Cardiopulmonary resuscitation during the COVID-19 pandemic in Spain, *Rev. Española Anestesiol. y Reanim. (English Ed.)*, Vol. 68, Issue 8, Oct. 2021, pp. 437-442.
- [7]. K. F. Sun, K. M. Poon, C. T. Lui, K. L. Tsui, Clinical prediction rule of termination of resuscitation for out-of-hospital cardiac arrest patient with pre-hospital defibrillation given, *Am. J. Emerg. Med.*, Vol. 50, 2021, pp. 733-738.
- [8]. G. D. Perkins, *et al.*, European Resuscitation Council guidelines for resuscitation 2015: Section 2. Adult

basic life support and automated external defibrillation, *Resuscitation*, Vol. 95, 2015, pp. 81-99.

- [9]. N. C. Chandra, *et al.*, Observations of ventilation during resuscitation in a canine model, *Circulation*, Vol. 90, Issue 6, 1994, pp. 3070-3075.
- [10]. T. W. Murphy, et al., Cardiac arrest: An interdisciplinary scoping review of the literature from 2019, *Resusc. Plus*, Vol. 4, 2020, 100037.
- [11]. I. Azcarate, *et al.*, The role of chest compressions on ventilation during advanced cardiopulmonary resuscitation, *J. Clin. Med.*, Vol. 12, Issue 21, 2023, 6918.
- [12]. J. J. Gutiérrez, et al., Standardisation facilitates reliable interpretation of ETCO₂ during manual cardiopulmonary resuscitation, *Resuscitation*, Vol. 200, 2024, 110259.
- [13]. M. Leturiondo, *et al.*, P111 Ventilation ratemay compromise clinical decisions based on ETCO₂ during CPR, *Resuscitation*, Vol. 175, 2022, S76.
- [14]. C. Sandroni, P. De Santis, S. D'Arrigo, Capnography during cardiac arrest, *Resuscitation*, Vol. 132, 2018, pp. 73-77.
- [15]. A. H. Idris, et al., High Incidence of Chest Compression Oscillations Associated with Capnography During Out-of-Hospital Cardiopulmonary Resuscitation, Am. Heart Assoc., 2010.
- [16]. M. Vanwulpen, M. Wolfskeil, C. Duchatelet, K. Monsieurs, S. H. Idrissi, Quantifying inspiratory volumes generated by manual chest compressions during resuscitation in the prehospital setting, *Resuscitation*, Vol. 118, 2017, e18.
- [17]. D. L. Grieco, et al., Intrathoracic airway closure impacts CO₂ signal and delivered ventilation during cardiopulmonary resuscitation, Am. J. Respir. Crit. Care Med., Vol. 199, Issue 6, 2019, pp. 728-737.
- [18]. A. I. Molina, W. J. Giraldo, J. Gallardo, M. A. Redondo, M. Ortega, G. García, CIAT-GUI: A MDE-compliant environment for developing Graphical User Interfaces of information systems, *Adv. Eng. Softw.*, Vol. 52, Oct. 2012, pp. 10-29.
- [19]. S. Ruiz de Gauna, et al., Characterization of mechanical properties of adult chests during pre-hospital manual chest compressions through a simple viscoelastic model, *Comput. Methods Programs Biomed.*, Vol. 242, 2023, 107847.
- [20]. Y. Tang, M. J. Turner, A. B. Baker, Systematic errors and susceptibility to noise of four methods for calculating anatomical dead space from the CO₂ expirogram, *Br. J. Anaesth.*, Vol. 98, Issue 6, Jun. 2007, pp. 828-834.
- [21]. O. Cornejo Zúñiga, R. Rebolledo Vega, Estimación de parámetros en modelos no lineales: algoritmos y aplicaciones, *Rev. EIA*, Vol. 25, 2016, pp. 81-98.
- [22]. Ó. C. Zúñiga, R. R. Vega, Estimation of parameters in nonlinear models: algorithms and applications, *Revista EIA*, Vol. 13, Issue 25, 2016, pp. 81-98.
(052)

Forecasting Flood in Vietnam Using Deep Learning

T. L. Nguyen¹ and T.H. Nguyen^{1,2}

¹Department of Informatics, Institute for Cybersecurity and Digital Technologies, MIREA – Russian Technological University ²University of Information and Communication Technology, Thai Nguyen University, 70000, Thai Nguyen, Viet Nam Tel.: +79648105759 E-mail: nguen@mirea.ru, thuhuongyb@gmail.com

Summary: This article presents an overview of deep learning applications to forecast the risk of flood disasters, aiming to minimize the damage caused by natural disasters. Natural disaster forecasting is a complex challenge, requiring the processing and analysis of large amounts of multi-source and multi-dimensional data. Recent studies have shown the superiority of machine learning models over traditional methods. Although there are many challenges, with the development of machine learning methods and digital data collection, we can build increasingly accurate models to forecast the risk of natural disasters, contributing to natural disaster prevention in Vietnam.

Keywords: Artificial intelligence, Deep learning, Convolutional neural network, Weather forecasting science, Flood warning system.

1. Introduction

Floods are an annual natural phenomenon that causes severe damage to infrastructure, crops and the economies of countries around the world. Forecasting the risk of natural disasters is an important task to prevent and minimize the damage caused by natural disasters. Conventional models still have many limitations, as they require a large number of input parameters such as water levels, flow rates, evaporation, infiltration rates, soil moisture, etc. and, above all, require a lot of time for simulation. In addition, the creation, simulation and analysis of the results of physically based models require the involvement of experts in hydrology, hydraulics and related fields. Therefore, the practical application of these models for real-time flood warning is still limited. A future direction for hydrology and water resource management is to find methods to integrate traditional mathematical models with machine learning models to directly process, analyze and extract information from big data sources. Therefore, machine learning has attracted the attention of hydrologists in recent years and has been applied in various fields due to its ability to process large amounts of data.

Deep learning is a technique that helps computers learn from data and can be applied to analyze the complex relationships between factors affecting natural disasters. Many studies on the application of machine learning in forecasting the risk of natural disasters have achieved remarkable results. Machine learning models have helped improve the accuracy and reliability of forecasting work in many places around the world. However, this work also poses many formulas, including choosing the appropriate influencing factors for the research area, the quality of the collected data, the quality of the machine learning model and practical application when conditions in the research area change. The main causes of floods are the impacts of climate change, human impact, and heavy rains that cause rapid increases in river water levels, making it difficult for water to drain [1]. Over the past 27 years, floods have caused the deaths of more than 175,000 people and caused severe economic impacts estimated at \$2.2 billion globally [2]. Flood response is very important, especially in developing countries, where disaster prevention and mitigation measures are limited and flood plains are often densely populated [3]. Therefore, flood forecasting, water levels and river flows, especially on rivers with few or no hydrological monitoring stations, are very important in warning people and local authorities of floods. There are many projects and studies by domestic and foreign scientists applying physicallybased models with high accuracy to predict water levels, flow or inflows on rivers, reservoirs, etc. However, these types of models still have many limitations because they require a lot of input parameters in the model such as water level, flow, evaporation, infiltration rate, soil moisture, etc. and especially need a lot of time to simulate. Moreover, to set up, simulate and analyze the output results of physically-based models requires the participation of experts in the fields of hydrology, hydraulics, etc. Therefore, the practical application of these models in flood warning over time is not high. The future direction of hydrology and water resources management is to find a way to integrate water resources management based on traditional mathematical models into machine learning models to directly process, analyze and extract information from large data sources [4].

Therefore, in recent years, machine learning has attracted much attention from hydrologists and has been widely applied in many fields thanks to its ability to manage large data. With the development of information technology, the terms machine learning or deep learning are no longer foreign to us. Deep learning used in many different professions and areas of society, including water management. The review articles have highlighted the development of research and the application of machine learning to problems in the field of water management and disaster risk management [5]. The current authors mostly focus more on explaining the algorithmic structure of machine learning rather than its applications, which makes it difficult to access for readers with a background in hydrology and water resources [6]. Deep learning used to predict the likelihood of flash floods [7, 8]. To build these models, information on geographical features, hydrometeorology, vegetation, and human activities in the study area used in the training data. This information includes elevation, slope, land morphology, rainfall, river flow, crop type, land use, resource exploitation activities, construction of structures, road construction, etc. at locations where flash floods have occurred in the past. The authors of the paper [9] presented a flood simulation. The accuracy of a data-driven model using machine learning in flood simulation evaluated by comparing it with traditional physics-based models. The study in [7] has given an overview of the strengths of some algorithms in machine learning in the problem of water level prediction. An overview of artificial intelligence (AI) models used in the field of flow prediction to contribute to the improvement and optimization in the management and operation of reservoirs proposed in the study [10-12].

2. Methodology

Convolutional neural network-CNN is an artificial neural network architecture that is mainly used in processing spatial data such as images and videos. However, in the CNN model for the problem of building flood maps, the data is transformed into a 2dimensional input matrix, where the first axis represents time and the second axis represents the feature variables. The "Convolutional" layers in CNN are used to extract local features from the time series data. The "Pooling" layers help reduce the spatial dimension of the extracted features, while the "Activation" layer creates nonlinearity in the model. Finally, the "Fully Connected" layers are used to predict the next value based on the extracted features (Fig. 1).

In deep learning problems in general and in flood forecasting problems in particular, simulating a machine learning model to predict the desired results will include 6 main steps. The steps to build a neural network include:

Step 1: Normalizing the input data and splitting the data set.

Step 2: Design of the network structure

Step 3: Initialization of random values for the weights.

Step 4: Perform the forward propagation phase through all layers of the network.

Step 5: Calculate the error on the initial learning set and decide whether to continuelearning or not. If you continue, go to step 5, otherwise exit the loop.

Step 6: Calculate the error at each node of each layer to update the weights and todetermine the constants.



Fig. 1. Feature maps of Flood warning system.

The dataset describes the flooding results simulated with a powerful integrated hydrodynamic modelling system for pluvial flooding and fluvial flooding in Vietnam. The pluvial flooding results simulated by HiPIMS are determined by the design rainfall in the 2, 5, 10, 20, 50 and 100 year return periods, and the

fluvial flooding results simulated by HiPIMS are determined by the river water level boundary in 2011 (Fig. 2) [13].



Fig. 2. Dataset of the flood.

The initial data set is split into two component data sets, the first data set is used as input data for the prediction problem, while the second data set is used to check the results of the prediction method. In the prediction problem with CNN, the first data set is fed into the network for training. After each iteration of the training process, the network saves the weight matrix and compares the error of the network with the initial allowable error. The result of the training process is that the network model has been trained so that the main data set is the weight matrix with the smallest error.

3. Analysis and Evaluate Experimental Results

Some challenges in building and using deep learning models to predict the risk of natural disasters are as follows:

Regarding data of influencing factors. This is a challenge related to the collection, pre-processing, selection and analysis of data related to natural disasters. The factors causing natural disasters are very diverse, from terrain, geology, weather to human impact and need to be selected appropriately for the research area. These data may be incomplete, inaccurate or inconsistent. The data needs to be preprocessed to remove noise, omissions, duplication and reformatted to suit the machine learning model.

Regarding models. This is a challenge related to the selection, training, testing and evaluation of machine learning models. There are many different types of machine learning models that can be used for forecasting the risk of natural disasters, but no model is optimal for all cases. Therefore, it is necessary to build different types of models and choose the best model. Deep learning models can be affected by problems such as overfitting, instability or lack of interpretability.

First is the CNN implemented from scratch and then second is using the VGG-16 pretrained network. CNN without using pretrained network gives 91.5% accuracy and with VGG-16 gave 95.7% accuracy (Fig. 3).

By using gate mechanisms to control the flow of information during the computation of hidden states, the LSTM model has significantly improved its ability to learn and understand long-term dependencies in sequential data compared to traditional RNN methods. These gates allow the LSTM to store important information from previous time steps and adjust the storage and transmission of information based on context. This capability enables the LSTM to effectively handle sequential data and model long-term dependencies within time series data. The CNN-LSTM model outperforms other AI models, with 94% of the predicted flow rate (Q) errors being below 0.05 m³/s, while the LSTM and DNN models have >98.7% and >91.4% of their predicted Q errors below 0.05 m³/s, respectively. This indicates that the CNN-LSTM model provides higher accuracy in predicting flow rates compared to the LSTM and DNN models.



Fig. 3. Accuracy of training and testing process, (a) CNN with pretrained, (b) CNN with VGG-16.

4. Conclusion

Forecasting the risk of flood disasters is an important and urgent task to minimize human and property losses, as well as protect the environment and biodiversity. Machine learning models have been widely and effectively applied in disaster forecasting, using multi-source and multi-dimensional data related to the factors causing and influencing natural disaster events. Recent studies have shown that machine learning models have the ability to outperform traditional methods in forecasting the risk of natural disasters. However, there are still many challenges and potentials to improve and develop machine learning models in this field. With the advancement of computer science and data collection technology, we can expect that machine learning models will play an increasingly important role in forecasting and

responding to natural disasters in the future. Future implementation and monitoring. This is a challenge related to the application and maintenance of machine learning models to predict the risk of natural disasters for the study area in the future when data on influencing factors may change. Deep learning models may no longer be accurate when encountering new or different conditions from the training conditions. For example, factors causing natural disasters may change seasonally, yearly or due to unusual climate events. Deep learning models need to be updated and adjusted over time to reflect the change in factors causing natural disasters.

References

- G. Bloschl et al., Changing climate both increases and decreases European river floods, *Nature*, Vol. 573, No. 7772, Sep. 2019, pp. 108–111.
- [2]. S. N. Jonkman, Global Perspectives on Loss of Human Life Caused by Floods, *Nat Hazards*, Vol. 34, No. 2, Feb. 2005, pp. 151–175.
- [3]. L. Alfieri, A global network for operational flood risk reduction, *Environmental Science & Policy*, Vol. 84, Jun. 2018, pp. 149–158.
- [4]. F. Ghobadi, D. Kang, Application of Machine Learning in Water Resources Management: A Systematic Literature Review, *Water*, Vol. 15, No. 4, Jan. 2023, 620.
- [5]. K. Kunverji, K. Shah, A flood prediction system developed using various machine learning algorithms, in *Proceedings of the 4th International Conference on Advances in Science* & *amp; Technology (ICAST2021)*, Mumbai, India, 7 May 2021.

- [6]. F. Y. Dtissibe, C. Titouna, Flood forecasting based onan artificial neural network scheme, *Nat. Hazards*, 104, 2020, pp. 1211–1237.
- [7]. D. Chitwatkulsiri, D. H. Miyamoto, Real-Time Urban Flood Forecasting Systems for Southeast Asia—A Review of Present Modelling and Its Future Prospects, Water, 15, 2023, pp. 178-185.
- [8]. V. Kumar, K. V., Sharma. Comprehensive overview offlood modeling approaches: A review of recent advances, *Hydrology*, 10, 2023, 141.
- [9]. H. M. Azamathulla, K.V. Sharma, The state of the art in deeplearning applications, challenges, and future prospects: A comprehensive review of flood forecasting and management, *Sustain ability*, 15, 2023, 10543.
- [10]. G. Furquim, B. S. Fical, Improving the accuracy of a flood forecasting model by means of machine learning and chaos theory: A case studyinvolving a real wireless sensor network deployment in Brazil, *NeuralComput. Appl.*, 27, 2016, pp. 1129–1141.
- [11]. M. P. Quang, K. Tallam, Predicting Flood Hazards in the Vietnam Central Region: An Artificial Neural Network Approach, *Sustainability*, 14, 19, 2022, 11861.
- [12]. H. D. Nguyen, D. K. Dang, Flood hazard assessment using machine learning and hydrodynamic modeling: case study in the Vu Ga–Thu Bon basin in Vietnam, *Water Practice* and Technology, 19, 10, 2024, pp. 4104–4127.
- [13]. Zhao, J.H.; Liang, Q.H., Flood modelling simulations for Can Tho city, Vietnam, NERC EDS Environmental Information Data Centre, 2022.

(059)

Enhancing Accuracy in Non-contact Physiological Monitoring: The Critical Role of Radar and Sensor Signal Alignment

Nour Ghadban, Mostafa Elsayed, Jonathan Cooper and Julien Le Kernec

¹ James Watt School of Engineering, University of Glasgow, Glasgow, UK Tel.: + 447442678815 E-mail: nour.ghadban@glasgow.ac.uk

Summary: Non-contact physiological monitoring utilizing radar technology has emerged as a promising approach for assessing heart rate (HR) and respiration rate (RR). However, misalignment between radar and reference sensor signals can introduce significant measurement errors, compromising the accuracy and reliability of radar-derived physiological parameters. This study investigates the critical role of signal alignment in Frequency Modulated Continuous Wave (FMCW) radar-based monitoring and proposes a cross-correlation technique to synchronize radar and sensor data. Experimental results demonstrate and emphasize the importance of signal alignment as a prerequisite for the clinical reliability of radar-based monitoring. Future research directions include the incorporation of artificial intelligence- driven techniques to enhance signal synchronization and improve robustness across diverse patient populations. By ensuring precise signal integration, radar-based physiological monitoring holds significant potential as a viable alternative to traditional contact- based systems, contributing to the broader efforts of making non- invasive health monitoring more accessible, reliable, and clinically validated for various

Keywords: Radar-based monitoring, Frequency Modulated Continuous Wave (FMCW) radar, Physiological signal processing, Cross-correlation alignment, Time-Gated Peak Detection (TGPD).

1. Introduction

healthcare applications.

The field of life sciences is seeing increased interest in non-invasive monitoring of vital signs like heart rate (HR) and respiration rate (RR), given its potential use in clinical and remote health surveillance [3, 4]. For non-contact health monitoring systems to be precise and dependable, it is essential to accurately align radar and sensor signals [5, 6].

While prior research has shown that Frequency Modulated Continuous Wave (FMCW) radar can detect HR and RR, obstacles remain in achieving precise signal alignment. Although radar technology has progressed, there is insufficient comprehensive research examining how signal misalignment impacts measurement accuracy in physiological monitoring. This investigation seeks to address this knowledge gap by exploring the consequences of misalignment between radar and sensor data and investigating methods to achieve accurate synchronization. What impact does the misalignment of radar and sensor signals have on the precision of HR and RR measurements in non-contact monitoring systems? This study aims to assess the efficacy of cross-correlation techniques in synchronizing radar and sensor signals to reduce measurement errors. We theorize that utilizing cross-correlation techniques for signal alignment will substantially decrease errors and improve the accuracy of radar-based HR and RR monitoring. The evolution of healthcare technology has enabled the creation of innovative non-contact monitoring systems for physiological parameters such as heart rate (HR) and respiration rate (RR). Traditional methods, including electrocardiography (ECG) and pulse oximetry, require direct physical contact, which can be invasive, uncomfortable, and impractical in certain medical environments. Radar-based monitoring offers a promising alternative by allowing remote and continuous assessment of vital signs.

2. Radar Technology in Healthcare

Radar technology, traditionally used in military and defense applications for detecting aircraft and ships, has seen significant adaptation for biomedical uses over the past few decades. This shift includes the development of commercially available radar systems, such as Frequency Modulated Continuous Wave (FMCW) radar operating between 2.4 GHz and 24 GHz, which can be employed for vital signs monitoring, including heart rate and respiration rate [1].

The fundamental mechanism of non-contact radar sensing involves emitting electromagnetic signals towards a subject. When these signals reflect off the body, the phase changes of the returned signal correlate directly with subtle movements, such as those caused by breathing and other cardiorespiratory activities [1].

This capability is enhanced by various signal processing techniques that extract vital signs from the reflected echoes, including breathing rate, heart rate, and tidal volume [1].

3. Cross-correlation Methodology

Cross-correlation is a fundamental technique utilized to align radar and sensor signals by identifying

the relationships between two signals at various time lags. It quantifies the degree of similarity between these signals by shifting one relative to the other and analyzing their overlap. This process facilitates the identification of patterns and time delays, which are essential for achieving accurate signal alignment. Mathematically, cross-correlation is defined as a function that quantifies the similarity between two signals, x[n] and y[n], based on a time lag m. It is expressed as (1):

$$\mathbf{R}_{\{xy\}}[\mathbf{m}] = \Sigma \mathbf{x}[\mathbf{n}] \cdot \mathbf{y}^*[\mathbf{n} \cdot \mathbf{m}], \qquad (1)$$

where y* represents the complex conjugate of y.

This function effectively quantifies the degree of correspondence between the signal y and a time-shifted version of the signal x. By computing this similarity measure, cross-correlation facilitates the detection of time delays and precise signal alignment, rendering it an essential tool in radar, sensor applications, and various signal processing domains [2].

4. Radar System Specifications

The radar system employed in this investigation operates with the following specifications, ensuring precise measurement and processing of physiological signals as presented in Table 1.

Parameter	Value
No. of Tx Channels	1
No. of Rx Channels	2
RF Centre Frequency	9.5
[GHz]	9.5
Radar Mode	FMCW
Bandwidth [GHz]	1
Chirp Duration [µs]	50
No. of Samples per Chirp	512
Doppler Processing	FD:BPF+FFT,TD: EEMD
Gold Standard	ECG, Respiration Belt
	(RB)

Table 1. Radar Settings for Measurements.

5. Methodology

This study was part of a larger project investigating physiological monitoring using radar and sensor technologies. Supported by the Engineering and Physical Sciences Research Council (EPSRC) – Quantum Imaging for Monitoring of Wellbeing and Disease in Communities (QUEST, EP/T021020/1) and the Scientific and Steering Committees of the Women and Science Chair. Conducted in compliance with the ethical standards of the University of Glasgow, it received approval on May 21, 2024, under reference number 300230110.Data were collected in a controlled laboratory environment. Five healthy participants were

recruited, each undergoing four repeated measurement sessions of one-minute duration. During each session, physiological signals were recorded using both radar and reference sensor devices. Participants were instructed to maintain a resting position under standardized conditions to minimize motion artifacts and external influences on the signals.

In this study, multiple signal processing algorithms were employed to extract and refine heart rate (HR) and respiration rate (RR) from radar and sensor data. The key steps included preprocessing, filtering, estimation, error calculation, and alignment using cross-correlation techniques. The following section provides a detailed breakdown of the algorithms utilized in each stage of processing.

5.1. Data Preprocessing

Prior to the application of signal processing techniques, radar and sensor data were imported from CSV files and parsed into structured time-series data. The extracted components comprised:

- Time (t): Represents the timestamp of each recorded sample;
- Heart signal (HR_signal): The raw physiological signal corresponding to heart rate;
- Respiration signal (RR_signal): The raw physiological signal corresponding to respiration rate.

5.1.1. Import CSV files containing radar and sensor measurements.

5.1.2. Extract time, HR, and RR components.

5.1.3. Compute sample spacing (Δt) to determine signal resolution.

5.1.4. Normalize signal amplitudes to maintain uniform scaling between radar and sensor signals.

5.2. Signal Filtering

To eliminate noise and unwanted frequency components, low-pass and bandpass filtering were applied separately for HR and RR signals.

5.2.1. Respiration Rate (RR) Filtering – Low-Pass Butterworth Filter.

- RR signals primarily reside within the 0.1–0.5 Hz frequency range;
- A 4th-order Butterworth low-pass filter was implemented to attenuate high-frequency noise.

5.2.1.1. Define cutoff frequency: 0.3 Hz.

5.2.1.2. Design a low-pass Butterworth filter utilizing scipy.signal.butter.

5.2.1.3. Apply the filter using scipy.signal.filtfilt for zero-phase distortion.

5.2.1.4. Extract peaks and troughs from the smoothed RR signal.

5.2.2. Heart Rate (HR) Filtering – FIR Bandpass Filter.

• HR signals typically occupy the 0.8–3.0 Hz frequency range;

- A finite impulse response (FIR) bandpass filter was employed to remove low-frequency drift and high-frequency artifacts.
- 5.2.2.1. Define passband frequencies: 0.8–3.0 Hz.

5.2.2.2. Design an FIR filter utilizing scipy.signal.firwin.

5.2.2.3. Apply filtering using scipy.signal.lfilter.

5.3. Physiological Parameter Estimation

5.3.1. Respiration Rate (RR) Estimation – Peak and Valley Detection.

- Respiration rate was derived through the identification of peaks and valleys in the respiration signal;
- The temporal difference between consecutive peaks was utilized to compute respiration rate.
- 5.3.1.1. Detect local maxima and minima in the RR signal using scipy.signal.find_peaks.

5.3.1.2. Compute the time difference (Δt) between successive peaks.

- 5.3.1.3. Convert the difference into breaths per minute (BPM).
- 5.3.2. Heart Rate (HR) Estimation Time-Gated Peak Detection (TGPD)
 - The Time-Gated Peak Detection (TGPD) algorithm was implemented to refine heart rate estimation.

5.3.2.1. Identify prominent peaks in the HR signal using scipy.signal.find_peaks.

5.3.2.2. Apply a time-gating threshold to filter out erroneous peaks.

5.3.2.3. Compute HR in beats per minute (BPM).

5.4. Error Calculation – Pre-alignment

Errors were computed to assess initial discrepancies prior to alignment.

Error Metrics Utilized:

- Mean Absolute Error (MAE);
- Root Mean Square Error (RMSE).

5.5. Signal Alignment – Cross-correlation

To minimize discrepancies, a cross-correlationbased alignment technique was implemented.

5.5.1. Compute the cross-correlation between radar and sensor signals.

5.5.2. Identify the lag k_{max} where correlation is maximized.

5.5.3. Shift the radar signal by k_max samples to align with the sensor signal.

5.5.4. Recalculate HR and RR post-alignment.

6. Results

Participants were recruited from a controlled laboratory setting. A total of five participants were

enrolled, with six participants completing the study. Data collection was successfully completed for all subjects, and no participants were excluded due to data inconsistencies or missing information. The study sample included 3 males and 2 females with an average age of 18-60 years (95 % CI: A–B). All participants were in good health, and their baseline physiological parameters were within normal ranges.



Fig. 1. Heart and Respiration Signal Extraction from Radar.

For this sample, we observe a temporal shift of 0.026 seconds (lag of 26 samples) in the cardiac signal, indicating a slight delay in its response. Conversely, the respiratory signal exhibits a temporal shift of -0.038 seconds (lag of -38 samples). Fig. 2 illustrates this relationship, demonstrating the shift differences between the two signals.

Following the application of cross-correlation and subsequent signal alignment, we observe a Respiration Signal Time Shift of 0.0 seconds and a Heart Signal Time Shift of 0.0 seconds. This observation indicates that both signals are now in perfect synchronization. Fig. 3 illustrates this alignment, demonstrating that the signals maintain phase coherence over the time interval from 0 to 0.6 seconds.

The accuracy of radar-derived HR, and RR measurements is assessed by comparing them to sensor data before alignment in Table 2.

 Table 2. Accuracy Assessment of Radar-Derived Heart

 Rate and Respiration Rate Measurements Before Alignment

 Compared to Sensor Data.

Object	HR (Radar)	RR (Radar)	HR (Sensor)	RR (Sensor)	Error (H)
1	70.18	18.32	86.32	16.69	18.697
2	64.54	14.61	83.71	25.48	22.900
3	63.79	19.51	72.59	23.26	12.122
4	66.07	22.59	82.61	26.34	20.021
5	68.17	15.87	88.60	16.28	23.058

After aligning the signals using cross-correlation, HR, and RR are recalculated to ensure an accurate comparison:

7. Discussion

This study demonstrates that implementing cross-correlation techniques for signal alignment significantly reduces errors in non-contact physiological monitoring using radar-based systems. Misalignment between radar and sensor signals can lead to inaccuracies in heart rate (HR) and respiration rate (RR) measurements. By applying cross-correlation-based alignment, the study achieved near-zero discrepancies, im- proving the reliability of radar-based monitoring. The study suggests that future research integrating AI-driven adaptive filtering could further enhance real-time synchronization, enabling widespread clinical adoption of radar-based monitoring in various healthcare applications.



Fig. 2. Time Shift and Lag Analysis of Heart and Respiration Signals.



Fig. 3. Aligned Heart and Respiration Signals After Cross-Correlation.

 Table 3. Accuracy Assessment of Radar-Derived Heart

 Rate and Respiration Rate Measurements After Alignment

 Compared to Sensor Data.

Object	HR (Radar)	RR (Radar)	HR (Sensor)	RR (Sensor)	Error (H)
1	70.18	18.32	86.32	16.69	18.697
2	64.54	14.61	83.71	25.48	22.900
3	63.79	19.51	72.59	23.26	12.122
4	66.07	22.59	82.61	26.34	20.021
5	68.17	15.87	88.60	16.28	23.058

Acknowledgments

This research was supported by the EPSRC – Quantum Imaging for Monitoring of Wellbeing and Disease in Communities, QUEST, EP/T021020/1 and the Scientific Committee and the Steering Committee of the Women and Science Chair.

References

[1]. M. Alizadeh, G. Shaker, J. C. M. De Almeida, P. P. Morita, S. Safavi-Naeini, Remote monitoring of

human vital signs using mm-wave FMCW radar, *IEEE Access*, Vol. 7, 2019, pp. 54958-54968.

- [2]. R. M. Burza, Overview of radar alignment methods and analysis of radar misalignment's impact on active safety and autonomous systems, *Sensors*, Vol. 24, Issue 15, 2024, 4913.
- [3]. R. Damaševičius, S. K. Jagatheesaperumal, J. Gorriz, et al., Deep learning for personalized health monitoring and prediction: A review, *Computational Intelligence*, Vol. 40, Issue 3, 2024, e12682.
- [4]. R. Damaševic'Ius, J. Gorriz, S. K. Jagatheesaperumal, et al., Deep learning for personalized health monitoring and prediction: a review, *Authorea*, June 20, 2023.
- [5]. M. Adeniyi, V. B. Ayoola, T. E. Samuel, W. Awosan, Artificial intelligence-driven wearable electronics and smart nanodevices for continuous cancer monitoring and enhanced diagnostic accuracy, *International Journal of Scientific Research and Modern Technology* (*IJSRMT*), Vol. 3, Issue 11, 2024, pp. 3-18.
- [6]. V. Sathyavathy, A novel approach to heart disease prediction using artificial intelligence techniques, *EAI Endorsed Transactions on Pervasive Health and Technology*, Vol. 10, 2024.

(060)

Radial Basis Operator Networks

J. A. Kurz¹, S. Oughton¹ and S. Liu²

¹ University of Waikato, School of Computing and Mathematical Sciences, Hamilton, Waikato 3216, NZ ² Clemson University, School of Mathematical and Statistical Sciences, Clemson, SC 29631, USA E-mail: jason.kurz@waikato.ac.nz

Summary: Operator networks are designed to approximate nonlinear operators, which map between infinite-dimensional spaces like function spaces. These networks are increasingly important in machine learning, particularly in scientific computing, due to their ability to handle data common in fields like climate modeling and fluid dynamics, where inputs are often discretized continuous fields (e.g., temperature or velocity distributions). We introduce the radial basis operator network (RBON), which represents a breakthrough as the first operator network capable of learning an operator in both the time and frequency domains when adjusted to accept complex-valued inputs. Despite the small, single hidden-layer structure, the RBON boasts small L^2 relative test error for both in- and out-of-distribution data (OOD) of less than 1×10^{-7} in some benchmark cases. Furthermore, it maintains small errors on OOD data from entirely different function classes than those used during training, showcasing its robustness and adaptability for advanced scientific applications.

Keywords: Operator networks, Neural operators, Radial basis functions, Machine learning, Scientific computing, Partial differential equations

1. Introduction

1.1. Background

Traditional feedforward neural networks (FNNs) and radial basis function (RBF) networks have been shown to be universal approximators of functions [1, 2], meaning they are capable of representing the mapping between finite dimensional spaces. Thus, these networks are limited in their design to predicting a measurement acting on a subspace of R^d for some $d \in \mathbb{Z}^+$. Operator networks, however, are designed to learn the mapping between infinite dimensional spaces; they receive functions as input and produce the corresponding output function. Scientific computing has benefited from using operator networks to enhance or replace numerical computation for the purpose of simulation and forecasting on a wide array of applications to include computational fluid dynamics and weather forecasting [3].

The two primary neural operators that demonstrated immediate success are the deep operator network (DeepONet) [4] based on the universal approximation theorem in [5], and the Fourier neural operator (FNO) The basic DeepONet [6]. approximates the operator by applying a weighted sum to the product of each of the transformed outputs from two FNN sub-networks. The upper sub-network, or branch net, is applied to the input functions while the lower trunk net is applied to the querying locations of the output function.

In contrast, the FNO is a particular type of Neural Operator network [7], which accepts only input functions (not querying locations for the output) and applies a global transformation on the function input via a more intricate architecture. Motivated by fundamental solutions to partial differential equations (PDEs), the FNO network sums the output of an integral kernel transformation to the input function with the output of a linear transformation. The sum is then passed through a non-linear activation function. To accelerate the integral kernel transformation, the FNO applies a Fourier transform (FT) to the input data, with the FT of the integral kernel assumed as trainable parameters.

Following their initial introduction, several extensions and modifications to FNO and DeepONet were introduced to improve performance in specific contexts. Examples include the Fourier-enhanced DeepONet [8] to improve DeepONet's robustness against Gaussian noise, U-FNO [9] and MIONet [10] introduce U-Net paths into the Fourier layer of the FNO architecture to improve accuracy for multi-phase flow applications, and model-parallel FNOs [11] parallelise the structure of FNO to reduce computation load for high-dimensional data. However, many of these situational improvements did not result in clear error reductions across a variety of contexts, at least not enough to justify the additional complexity in architecture contained in some of the proposed methods. This has changed with the recent introduction of a new neural operator.

The Laplace Neural Operator (LNO) [12] has recently become a benchmark standard for operator networks due to its improved handling of transient responses and non-periodic signals, limitations inherent in the Fourier Neural Operator (FNO). LNO achieves this by leveraging the pole-residue method to represent both transient and steady-state responses in the Laplace domain, leading to better test performance on out-of-distribution (OOD) data in most contexts. Additionally, LNO boasts a reduced training cost and a simpler network architecture. For these reasons, we have selected LNO as the primary comparison for our new operator network, alongside FNO and DeepONet. To thoroughly evaluate performance, we include a problem scenario from [12] that highlights LNO's small OOD error in predictions.

1.2. Our Contributions

We propose the radial basis operator network (RBON) based on the universal approximation theorem in [13]; a novel operator network that is, to the best of our knowledge, the first to be entirely represented with radial basis functions:

- The universal approximation result in [13] is extended to normalised RBONSs (NRBONs);
- The RBON is the first network to successfully learn an operator entirely in both the time domain and frequency domain, by altering the algorithm to accept complex data types;
- Despite the simple single-hidden-layer structure, the particular implementation of the RBON within demonstrates impressively small error on both in-distribution (ID) and OOD data, outperforming LNO by several orders of magnitude;
- The RBON demonstrates successful results on the first OOD example where the OOD input is an entirely different base function. Typically, OOD input functions for introducing new operator networks are a scaling, shifting, or simple transformation of the input functions used in training.

While operator networks are usually tested only using data generated from known systems, such as in systems of partial differential equations (PDEs), we include a scientific application where the data is real physical measurements and the underlying operator is unknown. This demonstrates the ability of RBON to make accurate forecasts for time-dependent systems, for the purposes of scientific experimentation. The rest of the paper is organised as follows, the theoretical foundation and details regarding the particular implementation are presented in Section 2, which precedes the results of the numerical experimentation first on generated data followed by the observed data in 3, with the discussion and conclusion at the end.

2. Methodology

The RBON is a numerical representation, G^{\dagger} , for an operator, $G: \mathcal{U} \rightarrow \mathcal{V}$, where \mathcal{U} and \mathcal{V} are infinite dimensional spaces, using radial basis functions. Following the work as shown in [13], we present, without proof, the universal approximation theorem for such a representation as well as extending the theorem to include NRBONs. The subsequent section details the precise implementation used for the experimental results.

2.1. Theoretical Foundation

In distribution theory the Schwartz space, $S(R^d)$, is the space of rapidly decaying functions that are infinitely differentiable and whose derivatives decay faster than a polynomial. Essentially, these are smooth functions that vanish quickly away from their center. The space containing all linear functionals that act on the Schwartz space is referred to as the space of tempered distributions and is represented symbolically as $S'(R^d)$; the prime notation connotes the duality relationship between the spaces. These spaces are for defining the necessary regularity for the radial basis functions used in the approximation.

Noting that C(A) represents all continuous functions defined on A, consider the functions g such that

$$g \in C(R) \cap S'(R), \tag{1}$$

meaning g is in the space of tempered distributions and is continuous on R. Choosing $||x||_{R^d}$ to represent the Euclidean norm for $x \in R^d$, we can represent a radial basis function acting on x as

$$g(\lambda \| x - \mu \|_{R^d}),$$

for constants $\lambda \in R, \mu \in R^d$. Then we have the following (see [13] for the proof with details).

Theorem 2.1. Suppose g is not an even polynomial and satisfies (1), X is a Banach space where $K_1 \subseteq X, K_2 \subseteq R^d$ are two compact sets in X and R^d respectively. Suppose also that U is a compact set in $C(K_1)$, G is a nonlinear continuous operator, mapping U into $C(K_2)$, then for any small positive ϵ , there are positive integers M, N, m, constants $\xi_i^k, \omega_k, \lambda_i \in R$, $k \in \{1, ..., N\}, i \in \{1, ..., M\}, m$ points $x_1, ..., x_m \in$ $K_1, c_1, ..., c_N \in \mathbb{R}^d$, such that

$$\left|G(u)(y) - G^{\dagger}(u^m)(y)\right| < \epsilon,$$

for every $u \in U$ and $y \in K_2$, where $u^m = (u(x_1), ..., u(x_m))$, and

for $\mu_{ik}^m = (\mu_{1k}^m, \dots, \mu_{mk}^m), k = 1, \dots, N.$ For ϵ and ξ_i^k given as in Theorem 2.1 set

$$\begin{split} \boldsymbol{\xi}_{i}^{k} &= \boldsymbol{\xi}_{i}^{k} \sum_{i=1}^{M} g(\lambda_{i} \| \boldsymbol{u}^{m} - \boldsymbol{\mu}_{ik}^{m} \|_{R^{m}}) g(\boldsymbol{\omega}_{k} \| \boldsymbol{y} - \boldsymbol{c}_{k} \|_{R^{d}}), \end{split}$$
(3)

and the corollary extending the theorem for the normalised representation follows immediately.

Corollary 2.1.1. Under the same assumptions in Theorem 2.1 and with ξ_i^k as defined in (3), we have

$$\left|G(u)(y) - \widetilde{G^{\dagger}}(u^m)(y)\right| < \epsilon,$$

where

$$\begin{aligned} \widetilde{G}^{\dagger}(u^m)(y) &= \\ &= \sum_{i=1}^M \sum_{k=1}^N \widetilde{\xi}_i^k \frac{g(\lambda_i \| u^m - \mu_{ik}^m \|_{R^m})g(\omega_k \| y - c_k \|_{R^d})}{\sum_{i=1}^M \sum_{k=1}^N g(\lambda_i \| u^m - \mu_{ik}^m \|_{R^m})g(\omega_k \| y - c_k \|_{R^d})} \end{aligned}$$

The RBON, as represented in (2), comprises two single-layer sub-networks of radial basis functions. This architecture extends the concept of RBF networks to operators, analogous to how DeepONet extended FNNs. The sub-network that processes the function input u^m is called the branch net. Here, u^m represents the input function u sampled at m point locations, as defined in the theorem. The trunk net, on the other hand, receives inputs corresponding to the domain locations where the network will produce output function values.

2.2. Practical Implementations

Having established the theoretical foundation, we now turn to the practical implementation of our approach. This section outlines the step-by-step process for the realised implementation of both RBON and NRBON. The implementation consists of several key steps that translate our theoretical model into a functional algorithm.

From Theorem 2.1, recall $u^m \in R^m$ represents the numerical approximation of the function u sampled at m locations, G^{\dagger} is the network approximation of the operator, G, mapping u^m to the function v at the query location $y \in R^d$. Then, given input functions u_j^m for $j \in \{1, ..., J\}$, and query locations y_1 for $l \in \{1, ..., L\}$ where J and L denote the number of training input functions and query points, respectively, we outline the process for finding the network parameters.

RBF transformations. In both the trunk and branch networks we employ Gaussian functions for the RBF transformations, defined as

$$\phi(x,c,\sigma) = \exp\left(-\frac{\|x-c\|^2}{2\sigma^2}\right),$$

where *c* and σ are the RBF centers and spreads. The RBF centers are determined using K-means clustering [14, 15] on the input data for each sub-network, with the spreads calculated based on inter-cluster distances. The branch and trunk network transformations on an input pair $\{u_j^m, y_l\}$, with *M* and *N* RBFs, are represented by the vectors

$$b(u_j^m) = \left[\phi(u_j^m, c_1^b, \sigma_1^b), \dots, \phi(u_j^m, c_M^b, \sigma_M^b) \right]^l, t(y_l) = \left[\phi(\mathbf{y}_l, c_1^t, \sigma_1^t), \dots, \phi(\mathbf{y}_l, c_N^t, \sigma_N^t) \right]^T,$$

where c_i^b , c_k^t are the RBF centers and σ_i^b , σ_k^t are spreads for the associated branch and trunk networks.

Weight parameter calculation. For each query location y_1 , we first compute

$$\Phi_{l} = [b(u_{1}^{m}) \otimes t(y_{l}), \dots, b(u_{l}^{m}) \otimes t(y_{l})],$$

where \otimes denotes the Kronecker product, making Φ_1 of dimension $NM \times J$. The weights ξ_1 of dimension $NM \times 1$, are then determined by solving

$$\xi_l^T \Phi_l = \left[v_1(y_l), \cdots, v_l(y_l) \right],$$

using the Moore-Penrose inverse [16-17]. This process yields *L* weight vectors ξ_1 , for each query point. The final weight vector ξ is obtained by element-wise averaging across the *L* vectors ξ_1 . Given the input u^m , the network approximation for the associated output function v at query point y is then

$$G^{\dagger}(u^{m})(y) = \mathcal{L}(\xi^{T}[b(u^{m}) \otimes t(y_{l})]),$$

where \mathcal{L} denotes a linear transformation applied to the final output whose parameters are solved for directly using the training data.

NRBON modification. The NRBON differs from RBON in normalizing the products of the branch and trunk outputs by dividing each element of the vector $[b(u^m) \otimes t(y_1)]$ by the vector's sum. This normalization adjusts the computation of Φ_1 by its column totals.

Using K-means to determine the parameter locations for the RBFs limits the number of RBFs in the representation by the size of the training data set. It is worth noting that manually assigning the centers for the RBFs produces satisfactory results, but tends to result in larger error than using K-means. Hence manually assigning centers is only advisable when working with small training sets. Moreover, the majority of the variation in train/test error is mostly due to the varying results from the location parameters determined by the K-means clustering.

Concluding the description of the practical implementation, we note that the network weights can be solved for using an iterative approach such as least-mean-squares, but results in weights that on average produce larger error in their predictions.

2.3. Learning in the Frequency Domain

The RBON is designed to learn the operator in the frequency domain as well as in the time domain. The frequency domain is a representation of signals or functions in terms of their frequency components, rather than time. It allows analysis of how signals vary with frequency, providing insight into characteristics like energy, power, and periodicity. The frequency domain is often used to examine cyclic behavior, separate overlapping signals, and simplify certain mathematical operations on signals. Considering that functions have a global representation in the frequency domain, this can have benefits in reducing the variability on the RBONs predictions for OOD data.

Thus, the RBON can be trained on functions in the time domain to approximate the operator G, or the Fourier transform, \mathcal{F} , can be used to convert functions to the frequency domain, in which case the RBON is learning the approximation in the frequency domain. This is especially beneficial for applications where the data is stored in the frequency domain representation.

3. Numerical Experiments and Results

This section is partitioned into *numerical computing* experiments and a scientific application based solely on data collected from observed measurements. This demonstrates the ability for the RBON to learn the mapping in a variety of contexts including when the mathematical representation for the operator is unknown. We define the numerical computing setting here as scenarios where the data is completely generated from numerical approximations of solutions to mathematical equations. Thus, the operator is known precisely and the results of the RBON can be compared to the numerical approximation of the operator output.

Alternatively, the governing equations for scientific applications are not always known and data is often aggregated from physical measurements. Distinguishing between settings using *generated* data as opposed to *observed* data shows the flexibility of the RBON and its ability to support scientific experimentation and forecasting.

For all the numerical computing experiments, we limited the size of the trunk and branch networks to be no larger than 15 nodes each, capping the number of multiplier parameters in the hidden layer at 225. These restrictions demonstrate the network's ability to maintain small errors even under incredibly strict size constraints. All code for the RBON learning representation was implemented using the Julia chosen for programming language [18], its high-performance numerical computing capabilities, and has been made available at https://github.com/jkurz119/RBON.

3.1. Numerical Computing

In each of the following settings, there is a governing system of PDEs defined on a spatio-temporal domain, $\Omega \equiv (0,T) \times (0,L)$ for some final time T > 0 and length L > 0. The operator network, G^{\dagger} , was trained to learn an operator *G* within the PDE framework that maps functions representing the initial state or forcing term to the solution over the entire domain.

The input functions for the network for the in distribution data will thus be a family of functions representing an initial state (or forcing term) and parameterized across a specified range of values. ID data was segmented to produce a validation and test set. The validation set was used to optimized the size of the network over a few selected options. The test set provides the in-distribution test error with the out-ofdistribution errors based on a set of input functions that have been more significantly altered from the in distribution data.

Wave Equation. Consider the wave equation of the form

$$\frac{\partial^2 u}{\partial t^2} = c^2 \frac{\partial^2 u}{\partial x^2} \text{ for } (t, x) \in \Omega,$$

where c is the speed of propagation of the wave and u(t, x) models the displacement of a string with Dirichlet boundary conditions. The operator network, G^{\dagger} , was trained to learn the mapping, G, from the initial state to the solution, $G: u_0(x) \rightarrow u(x, t)$. For the ID data, we particularize the initial condition as

$$u_0(x) = 2e^{-(x-\frac{L}{2})^2} + \frac{ax}{L},$$

where *a* is parameterized across the range [1,4] with step size 0.001. The OOD test set uses the same base function for $u_0(x)$, but for values of *a* in the range [5,5.5].

Burgers Equation. Consider the well-known Burgers' equation

$$\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} = v \frac{\partial^2 u}{\partial x^2}$$
, for $(t, x) \in \Omega$,

subject to homogeneous Dirichlet boundary conditions and under the following initial conditions $u_0(x) = a \sin \pi x$ where *a* ranges across the interval [0.1,5]. The operator learned is thus $G: u_0(x) \rightarrow u(x, t)$ and the RBON is tested on the set of polynomial functions $u_0(x) = bx(x - 1)$ where *b* is in the range [3.5,4.5] for the OOD data. Successful testing on the polynomial input after only being trained on the sine function is quite remarkable. The numerical data was generated using the exact solutions as derived in [19].

Euler-Bernoulli Beam Equation. The Euler-Lagrange equation for a dynamic Euler-Bernoulli homogeneous beam is

$$EI\frac{\partial^4 u}{\partial x^4} + \rho A\frac{\partial^2 u}{\partial t^2} = f(t,x) \text{ for } (t,x) \in \Omega,$$

where *E* is Young's modulus, *I* is the second moment of the area of the beam's cross section. The beam's density is denoted by ρ and the cross-sectional area as *A*. The operator network learns the mapping from the source term to the solution: $G: f(t, x) \rightarrow u(t, x)$. For the ID data we particularize the source term as $f(t,x) = ae^{-0.05x}(1-10^2)\sin(10t)$ for *a* in [0.05,10]. The source term for the out of distribution data is $f(t,x) = ae^{-x}(1-10^2)\sin(10t)$ for *a* in [1.24,10.19]. This scenario was used for testing in [12]. We include it here for benchmarking purposes, but note we *decrease* the size of the training set to include in-distribution test error.

Results. The results from these experiments can be seen in Table 1, which displays the average L^2 error for each function in the ID and OOD test sets. The L^2 relative error is computed as follows

$$L^2 \text{ rel. } error = \frac{\left| |v^{true} - v^{pred}| \right|_2}{\left| |v^{true}| \right|_2},$$

where v^{true} represents the ground truth values at all query points in the space-time domain, and v^{pred}

represents the network's predictions at the same locations.

The margin of error (MOE), shown in parentheses, was computed by multiplying the standard error of the point estimate by the critical value for a 95 % confidence level. While the strict interpretation of the confidence interval is limited due to the non-independence of observations, the MOE still provides an indication of variability in the average error.

Across the majority of problems, the RBON variants outperform the LNO, with the NRBON achieving consistently superior performance on both ID and OOD data. Overall, RBON variants collectively tend to outperform other operator networks, with one exception. Notably, operator networks generally exhibit smaller errors for the Beam equation due to their ability to accurately represent

linear operators. Networks that leverage global representations – such as FNO, LNO, and F-RBON (which trains on data with global representations) – tend to generalize better, while other networks overfit ID data and have significantly worse performance on OOD data. This difference is especially noticeable with the OOD input data for the Wave problem due to its highly oscillatory behavior.

DeepONet initially suffered from overfitting, resulting in poor OOD performance, but early stopping significantly improved its OOD errors, albeit at the cost of slightly worse ID errors. However, this improvement came at the expense of efficiency: DeepONet required significantly larger sub-networks, with over 10,000 products between trunk and branch outputs, compared to fewer than 200 products in the RBONs.

Network	In/Out	Wave	Burgers	Beam
DDON	In	9.4E-4(4.9E-5)	3.6E-3(6.0E-4)	4.1E-8(3.3E-6)
KDUN	Out	1.0E-1(2.0E-3)	2.6E-1(1.3E-2)	1.5E-1(2.5E-7)
NDDON	In	1.2E-5(9.4E-7)	3.3E-3(9.0E-4)	1.6E-7(2.4E-7)
INKBUN	Out	3.2E-1(1.1E-2)	1.0E - 1(5.7E - 3)	2.0E-8(4.9E-9)
E DRON	In	3.0E-6(2.2E-7)	5.9E-3(1.1E-3)	1.1E-1(1.3E-1)
r-KDUN	Out	8.6E-3(1.7E-4)	2.3E-2(5.5E-3)	6.6E-2(7.0E-3)
LNO	In	5.6E-1(1.1E-3)	1.7E-1(4.3E-4)	1.0E-2(3.9E-3)
LNU	Out	5.9E-1(9.2E-4)	2.0E-1(8.0E-6)	6.8E-3(1.5E-3)
ENO	In	9.9E-4(2.3E-5)	9.3E-3(1.2E-3)	4.0E-3(6.1E-3)
rnu	Out	1.1E-1(1.4E-3)	1.7E-2(7.0E-6)	1.5E-3(2.2E-4)
DON	In	5.3E-2(2.5E-4)	9.9E-1(4.0E-5)	2.9E-1(2.9E-1)
DON	Out	4.9E-2(3.4E-5)	9.9E-1(2.0E-6)	2.5E-1(1.4E-2)

Table 1. Average relative L^2 error on ID/OOD test data reported with margin of error in parentheses.

3.2. Scientific Application

Modeling the relationship between atmospheric CO_2 and global temperature is a complex process involving a large number of variables with many of them potentially unknown [20]. Focusing specifically on an operator that does not have a well-defined mathematical representation, we demonstrate the capacity of the RBON to learn the mapping between monthly atmospheric CO_2 measurements and both local and average global monthly temperatures. This provides a template for prediction and forecasting with the RBON based on collected data.

For this section, the RBON is used to learn the operators

$$G_{avg}: u(t) \to T^{avg}(t), \\ G_{loc}: u(t) \to T^{loc}(t),$$

where u represents the atmospheric CO₂ defined for tin a given time interval, and T^{avg} represents the average global temperature as published in [21]. The function T^{loc} is local temperature readings at the same site location where the CO₂ data was collected. Specifically, atmospheric CO₂ concentrations (ppm) derived from in situ air measurements at the well known Mauna Loa, Observatory, Hawaii [22]. The local temperature readings are much more variable than the global average and hence less easily predicted.

The nature of the operators G_{avg} and G_{loc} is expected to evolve in time due to fluctuations in other contributing factors, however, when continuously updating the RBON with new data, the predictions become quite accurate. While, it is possible to feed the CO_2 readings into the network as one function, the centers for the RBFs must be set manually as the K-means algorithm requires at least two function inputs. Instead, it is preferable to parameterize the functions across the years such that $t \in \{1, 2, ..., 12\}$ with each number corresponding to the month of the year the measurement was taken. Then operators are thus more accurately represented as

$$\begin{split} G_{avg} &: u_n(t) \to T_n^{avg}(t), \\ G_{loc} &: u_n(t) \to T_n^{loc}(t), \end{split}$$

where n corresponds to a specific year. Training on the historical data, omitting years with incomplete data, yields remarkable accuracy in the RBON predictions as shown in Table 2.

Results. The results in Table 2 highlight the effectiveness of RBON in accurately predicting both local monthly average temperatures and global average temperatures. To evaluate the forecasting accuracy, we trained RBON and NRBON networks on historical temperature data, withholding the most recent two or five years from the training set for testing. In addition to these models, we compared their performance against LNO, DeepONet, FNO, and LSTM. This comprehensive evaluation demonstrates the robustness of RBON across diverse benchmarks, including traditional time-series approaches such as LSTM [23] and as well as other operator networks.

Based solely on monthly CO_2 measurements and the month encoding for querying the output temperature, the RBON maintains an L^2 relative error of less than 10 %, with NRBON performing similarly.

Fig. 1 displays a comparison between the trained RBON networks' global temperature predictions and actual global temperature readings. The left graph shows the results when holding out the most recent two years, while the right graph illustrates the outcome when holding out the most recent five years of data. Interestingly, several networks - including RBON, F-RBON, DeepONet and LSTM - performed similarly on the smoother global temperature data. However, performance on the more variable local temperatures at the observatory publishing the atmospheric CO₂ measurements [22] provided a clearer distinction as RBON outperformed other networks, which struggled to capture the finer-scale variations in the data. Fig. 2 provides the visual comparison for local temperatures versus the predictions from the RBON variants. Note that temperature data for the local set was only available through 2018.

Table 2. Average rel. L^2 test error on local temp. Data.

Global temp:							
2 yr	0.02	0.14	0.02	0.96	0.02	0.31	0.01
5 yr	0.02	0.15	0.01	0.97	0.02	0.44	0.01
Local temp:							
2 yr	0.07	0.13	0.04	0.94	0.35	0.18	0.15
5 yr	0.07	0.13	0.13	0.95	0.51	0.22	0.14



Fig. 1. Two (left) and five (right) year global temperature predictions based on CO₂ input. Forecast values in shaded region.



Fig. 2. Two (left) and five (right) year local temperature predictions based on CO input. Forecast values in shaded region.

The significance of this result implies a robust model capable of providing reliable future temperature projections based on various atmospheric CO_2 scenarios under different climate responses. This robustness stems from the model's ability to isolate the impact of CO_2 on temperature, as the effects of other contributing elements are learned in the operator approximation. While predicting solely based on CO_2 measurements provides a simple example, there is an opportunity to include other contributing factors in the operator input to understand how co-variation among several input variables may affect the output.

Testing revealed that increasing the width of the branch and trunk networks enhances the model's flexibility to match highly variable and erratic behavior. However, given highly oscillatory data, the plain RBON can occasionally produce peaks and valleys that deviate too far from the data range when increasing model width. In contrast, the NRBON can increase its network size without generating extreme peaks. Consequently, the smaller RBONs used yield a more stable regression appearance, while the larger NRBON networks produce outputs that attempt to capture more of the random extreme values. This results in a slightly higher error (≤ 0.17) for the NRBON, but a shape that more closely resembles the true graph.

For completeness, we include all results pertaining to learning the operator in the frequency domain, namely the F-RBON. These results are presented in Table 2. It's worth noting that this dataset does not naturally lend itself to a Fourier transform, and the additional computational work is unnecessary since the representation in the time domain is sufficient.

4. Discussion and Conclusion

The RBON and its variants offer a simple yet powerful network architecture with prediction capabilities that yield errors smaller than the current leading operator network. The network's compact size provides opportunities for enhanced interpretability and reduced computational load, allowing for exact solutions of network parameters. Most variation across training cycles arises from the location and scale parameters of the RBFs, largely due to K-means' tendency to converge on local extrema. This variability can lead to errors differing by several orders of magnitude between runs of the K-means algorithm. A practical solution is to run K-means multiple times and select the configuration that minimizes the overall within-cluster distances. Furthermore, the RBON serves as an excellent tool for scientific computing, where recent advancements have only begun to explore the potential of operator networks in various fields. Finally, the RBON's ability to train on both real and complex-valued inputs, combined with its other strengths, makes it a promising candidate for applications in signal processing and computer vision tasks. The results in Table 2 demonstrate that RBON achieves superior accuracy across all PDE

benchmarks, maintaining robustness even for OOD test cases.

References

- K. Hornik, M. Stinchcombe, H. White, Multilayer feedforward networks are universal approximators, *Neural Networks*, Vol. 2, Issue 5, 1989, pp. 359-366.
- [2]. J. Park, I. W. Sandberg, Universal approximation using radial-basis-function networks, *Neural Computation*, Vol. 3, Issue 2, 1991, pp. 246-257.
- [3]. K. Azizzadenesheli, N. Kovachki, Z. Li, M. Liu-Schiaffini, J. Kossaifi, A. Anandkumar, Neural operators for accelerating scientific simulations and design, *arXiv preprint*, 2023, arXiv:2309.15325.
- [4]. L. Lu, P. Jin, G. Pang, Z. Zhang, G. E. Karniadakis, Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators, *Nature Machine Intelligence*, Vol. 3, 2021, pp. 218-229.
- [5]. T. Chen, H. Chen, Universal approximation to nonlinear operators by neural networks with arbitrary activation functions and its application to dynamical systems, *IEEE Transactions on Neural Networks*, Vol. 6, Issue 4, 1995, pp. 911-917.
- [6]. Z. Li, N. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. Stuart, A. Anandkumar, Fourier neural operator for parametric partial differential equations, in *Proceedings of the International Conference on Learning Representations*, 2021.
- [7]. N. B. Kovachki, Z.-Y. Li, B. Liu, K. Azizzadenesheli, K. Bhattacharya, A. M. Stuart, A. Anandkumar, Neural operator: Learning maps between function spaces with applications to PDEs, *J. Mach. Learn. Res.*, Vol. 24, 2023, pp. 89:1-89:97.
- [8]. M. Zhu, S. Feng, Y. Lin, L. Lu, Fourier-DeepONet: Fourier-enhanced deep operator networks for full waveform inversion with improved accuracy, generalizability, and robustness, *Computer Methods in Applied Mechanics and Engineering*, Vol. 416, 2023, 116300.
- [9]. G. Wen, Z. Li, K. Azizzadenesheli, A. Anandkumar, S. M. Benson, U-FNO – an enhanced Fourier neural operator-based deep-learning model for multiphase flow, *Advances in Water Resources*, Vol. 163, 2022, 104180.
- [10]. Z. Jiang, M. Zhu, L. Lu, Fourier-MiONet: Fourier-enhanced multiple-input neural operators for multiphase modeling of geological carbon sequestration, *Reliability Engineering & System Safety*, Vol. 251, 2024, 110392.
- [11]. T. J. Grady, R. Khan, M. Louboutin, Z. Yin, P. A. Witte, R. Chandra, R. J. Hewett, F. J. Herrmann, Model-parallel Fourier neural operators as learned surrogates for large-scale parametric PDEs, *Computers* & *Geosciences*, Vol. 178, 2023, 105402.
- [12]. Q. Cao, S. Goswami, G. E. Karniadakis, Laplace neural operator for solving differential equations, *Nature Machine Intelligence*, Vol. 6, 2024, pp. 631-640.
- [13]. T. Chen, H. Chen, Approximation capability to functions of several variables, nonlinear functionals, and operators by radial basis function neural networks, *IEEE Transactions on Neural Networks*, Vol. 6, Issue 4, 1995, pp. 904-910.

- [14]. S. Lloyd, Least squares quantization in pcm, *IEEE Transactions on Information Theory*, Vol. 28, Issue 2, 1982, pp. 129-137.
- [15]. E. W. Forgy, Cluster analysis of multivariate data: efficiency versus interpretability of classifications, *Biometrics*, Vol. 21, 1965, pp. 768-769.
- [16]. E. H. Moore, On the reciprocal of the general algebraic matrix, *Bulletin of the American Mathematical Society*, Vol. 26, Issue 9, 1920, pp. 394-395.
- [17]. R. Penrose, A generalized inverse for matrices, Mathematical Proceedings of the Cambridge Philosophical Society, Vol. 51, 1955, pp. 406-413.
- [18]. J. Bezanson, A. Edelman, S. Karpinski, V. B. Shah, Julia: A fresh approach to numerical computing, *SIAM Review*, Vol. 59, Issue 1, 2017, pp. 65-98.
- [19]. T. Öziş, E.N. Aksan, A. Özdeş, A finite element approach for solution of burgers' equation, *Applied Mathematics and Computation*, Vol. 139, Issue 2, 2003, pp. 417-428.

- [20]. B. J. Mills, A. J. Krause, C. R. Scotese, D. J. Hill, G. A. Shields, T. M. Lenton, Modelling the long-term carbon cycle, atmospheric CO₂, and Earth surface temperature from lateneoproterozoic to present day, *Gondwana Research*, Vol. 67, 2019, pp. 172-186.
- [21]. Our World in Data, https://ourworldindata.org/ grapher/monthly-average-surface-temperatures-byyear
- [22]. C. D. Keeling, S. C. Piper, R. B. Bacastow, M. Wahlen, T. P. Whorf, M. Heimann, H. A. Meijer, Exchanges of Atmospheric CO₂ and 13CO₂ with the Terrestrial Biosphere and Oceans from 1978 to 2000, I. Global Aspects, SIO Reference Series, 01-06, *Scripps Institution of Oceanography*, San Diego, 2001.
- [23]. S. Hochreiter, J. Schmidhuber, Long short-term memory, *Neural Computation*, Vol. 9, Issue 8, 1997, pp. 1735-1780.

(062)

The Protocol for Integration of Automated and Dynamic Facial Expression Emotion Recognition with EEG for Emotional Traits Analysis in Pilot Candidates

<u>S. Michalak</u>^{1,2}, T. Łodygowski³, P. Śniatała³, M. Goralewski², E. Kozielewska-Zwierska²,
 J. Moskal², M. Galant-Gołębiewska³, M. Maciejewska³, K. Śniatała³ and P. Zych³
 ¹ Poznan University of Medical Sciences, Przybyszewskiego 49, 60-355 Poznan, Poland,
 ² Institute of Neurological Disorders, Przybyszewskiego 49, 60-355 Poznan, Poland

³ Poznan University of Technology, Piotrowo 2, 60-965 Poznan, Poland,

E-mail: swami@ump.edu.pl

Summary: Pilots face unique psychological challenges and possess distinctive psychological traits. Research in aviation psychology has shown that pilots exhibit increased levels of assertiveness, activity, and a propensity for excitement-seeking. Initially, reactions were analyzed manually using the Facial Action Coding System (FACS), which includes 46 action units (AUs). Today, automated systems based on artificial intelligence, like FaceReader (by Noldus), have been developed. Other physiological parameters, such as blood pressure, heart rate, heart rate variability (HRV), and skin conductance, provide additional methods for recognizing emotions. The electroencephalogram (EEG) allows for a deeper understanding of emotions by examining bioelectrical responses and cognitive appraisal processes. This paper presents a protocol for the integration of automated and dynamic facial expression emotion recognition with EEG for the analysis of emotional traits in pilot candidates. The system is tested at the Poznan University of Technology Aviation Training Center. Additionally, the collected data will be used to build an AI model, which is intended to be used to support personalized neurorehabilitations.

Keywords: Emotion recognition, EEG analysis, Neurological evaluation.

1. Introduction

Pilots face unique psychological challenges and possess distinctive psychological traits. Research in aviation psychology has shown that pilots exhibit increased levels of assertiveness, activity, and a propensity for excitement-seeking [1]. Additionally, they demonstrate a more effective capacity to manage fear, anxiety, and stress compared to the general population [2, 3]. Studies using questionnaires [4] have established a positive correlation between global emotional intelligence (EI) scores and safety citizenship (SC), suggesting that EI can be a predictor of SC. Furthermore, EI is positively associated with safety behaviors, leading to enhanced safety standards. Emotional intelligence encompasses the ability to recognize, understand, and manage one's own emotions, as well as those of others. Facial expression analysis offers a way to monitor emotional states and reactions through the examination of facial muscle activity. Initially, reactions were analyzed manually using the Facial Action Coding System (FACS), which includes 46 action units (AUs). Today, automated systems based on artificial intelligence, like FaceReader (by Noldus), have been developed. Other physiological parameters, such as blood pressure, heart rate, heart rate variability (HRV), and skin conductance, provide additional methods for recognizing emotions. The electroencephalogram (EEG) allows for a deeper understanding of emotions by examining bioelectrical responses and cognitive appraisal processes.

1.1. EEG in Pilots Testing

The main interest of researchers in the assessment of the psychophysical state of pilots is the measurement of task load. This measurement aims not only to determine the current level of fatigue and stress, but also to predict their impact on the pilot's performance during long-term flight operations. In this complex system of interdependencies, the key role is played by the human ability to adapt to dynamically changing conditions. Task load directly affects the efficiency and correctness of the pilot's work [5]. EEG allows to determine the mental state of the person being examined. The results are most reliable and can provide correct information if the test is conducted in laboratory conditions. In the real working environment of a pilot (cockpit), there are too many external factors such as machine and instrument movements, current intensity, which can disturb the signal measured by the electrodes [6]. Monitoring the psychophysical condition of drivers can contribute to increased road safety. Regular use of such methods in transport practice allows for early detection of fatigue and implementation of preventive measures. Many specialists have attempted to link the results of electroencephalography tests with the level of task load imposed on the subjects. Signals measured by EEG strongly correlate with task load states, which has been confirmed many times by analyzing EEG measured while the subjects were performing demanding tasks [7, 8]. In February 2022, scientists from France conducted research that showed that the most sensitive to increased task load is the theta wave

band, the readings of which decreased with this increase. This was particularly visible in the frontal lobe area. On the other hand, alpha wave activity decreased. In the beta wave band, activity increased, but this was a moderate trend [9]. Other studies were conducted in Canada in 2023. They showed that beta and theta wave activity took on different trends depending on the physical load given to the subject with medium physical load, activity in their bands increased slightly, while with high physical load it decreased. Delta waves, on the other hand, increased in each case if they were accompanied by an increase in task load [6]. Many other studies show that as the difficulty of a task increases, activity in the frontal lobe of the brain decreases, while theta activity increases [10]. A study conducted in China in 2024 focused on measuring brain waves during a turning maneuver, which involves a greater workload for the pilot. It was concluded that during the turning phase, there is a significant increase in the amplitudes of beta waves, and a decrease in delta, theta and alpha waves [11]. Similar conclusions were drawn from studies in Chengdu, China, which showed that a decrease in alpha frequency band power in the temporal lobe corresponds to high awareness and deep thinking [12]. Scientists W. Lang and A. Mecklinger also showed that alpha waves decrease the more complex the task the subject has to perform [13]. In most studies, it is indicated that the increase in task load is accompanied by: a decrease in alpha waves, an increase in beta waves, an increase in theta waves. Gamma waves are very rarely discussed, and the relationships of delta waves are ambiguous and differ for different studies.

2. Proposed Protocol

We propose a protocol, presented in Fig. 1, for integrating automated and dynamic facial expression emotion recognition with EEG to analyze emotional traits in pilot candidates. The protocol involves three parts: text exposure, aviation simulator training, and a questionnaire. Candidates will be exposed to various texts: glossolalic recitations, metaphoric texts, children's poem, and plain texts-separated by 1-minute intervals of white noise (50 Hz) to establish baseline reactions as presented in Fig. 2. The FaceReader software will analyze video recordings (Noldus, Netherlands) to detect emotions like happiness, sadness, surprise, anxiety, anger, and disgust based on facial AUs. The glossolalic text consists of speech-like sounds that lack meaning, while a metaphor uses one word or phrase in place of another to suggest a likeness or analogy, activating the right insula, left temporal pole, and right inferior frontal gyrus (Schmidt, 2009). Simultaneously, EEG traces will be recorded using the Unicorn System (g.tec), along with physiological parameters including ECG, heart rate, and galvanometric skin conductance measurements. Next, the same data - comprising facial expressions, EEG traces, and physiological parameters - will be collected during the aviation simulation

training. Finally, candidates will complete the Trait Emotional Intelligence Questionnaire-Short Form, which consists of 30 items designed to measure overall trait emotional intelligence. Moreover, Hamilton Anxiety Rating Scale (HAM-A) will be applied as a measure of psychic anxiety (mental agitation and psychological distress) and somatic anxiety (physical complaints related to anxiety) [14]. Afterward, the collected data will be integrated and analyzed using taxonomy statistical methods and machine learning to create an emotional profile of the pilot candidates.

2.1. Data Analysis Procedures

Analysis of electroencephalogram (EEG) signals using computer algorithms is a field that has been of interest to scientists for a long time. Among the research areas discussed in the literature classification of disease conditions and emotions recognition may be distinguished. In particular, in terms of the classification of disease conditions, the following may be indicated:

- depression recognition [15, 16];
- epilepsy diagnosis and epileptic seizure focus detection [17-19];
- Parkinson's Disease detection [20, 21],
- mental disorders detection and prediction [22, 23];
- neonatal EEG interpretation support [24].

Studies focusing on the recognition of emotions based on the analysis of EEG signals can also be found. For instance, the study [25] focuses on the development of an emotion recognition model based on the analysis of EEG signals in order to assess the quality and user satisfaction with the product, while the study [26] uses the emotion recognition mechanism to assess the effect of public art psychotherapy and determine the public's evaluation of public art. The paper [27] reviews the possibilities of using generative AI for emotion recognition. Another application of EEG analysis that is gaining popularity may be motor imagery. Motor imagery may be understood as performing a movement on a mental level, without the participation of muscles, and which may lead to the activation of the same areas of the brain as the actual performance of such movements [28]. Among the papers reviews this topic, the studies [29, 30] may be indicated. In addition, among the topics considered by researchers in the presented scope, the analysis of cognitive load and attention detection may be pointed out. For instance, the reviews [31-33] are focused on this research topic. Among the classes of algorithms used in the automatic analysis of EEG signals, Machine Learning (ML) and Deep Learning (DL) methods are gaining popularity. For instance, the paper [34] reviews the classes of supervised classification algorithms used, reporting Support Vector Machine (SVM), Convolutional Neural Network (CNN), k-nearest neighbors (KNN), and Random Forest (RF) as some of the potentially accurate ones for classification tasks in the discussed scope.



Fig. 1. The proposed protocol design.



Fig. 2. Text exposure sequence used in the test.

2.2. Results Medical Interpretation

The proposed protocol enables investigation of relationship between objectively evaluated emotional trait and safe performance of pilot candidate. Due to exposition to stimuli incorporated in the protocol the adaptability of candidates and adaptive designs can be identified. Anxiety behaviors which lead to difficulty in concentration cause negative changes to both the psychomotor and attention skills in dynamic tasks related to aviation. Thus, identification of the tendency towards anxiety enables the application of cognitive and behavioral treatments for anxiety by the application of mindfulness practices like Mindfulness-Based Stress Reduction and Mindfulness-Based Cognitive Therapy.

3. Conclusion

We propose a protocol (Fig. 1) for integrating automated and objective analysis of emotional trait with the identification of adaptability design and anxiety behavior. Next, the revealed profile will be used for modification of pilot training, and when needed to the application of therapies Currently, the developed test bench is undergoing validation and minor modifications are being made to improve the test run. In parallel, appropriate computer tools are being developed, using NN to analyze EEG waveforms and aggregate the collected data.

References

- [1]. R. L. Grice, L. C. Katz, Personality Profiles of US Army Initial Entry Rotary Wing Students Versus Career Aviators, U.S. Army Research Institute for the Behavioral and Social Sciences, 2007.
- [2]. J. D. Callister, R. E. King, P. D. Retzlaff, R. W. Marsh, Revised NEO personality inventory profiles of male and female US Air Force pilots, *Mil. Med.*, 1999, Vol. 164, Issue 12, pp. 885-890.
- [3]. Y. Gao, S. Kong, Personality types of pilot students: A study of an Australian collegiate aviation program, *Int. J. Aviat. Aeronaut. Aerosp.*, Vol. 3, Issue 3, 2016, 6.
- [4]. Z. T. Dugger, B. McCrory Emotional intelligence and safety citizenship among army aviator, *Int. J. Aviat. Aeronaut. Aerosp.*, Vol. 8, Issue 1, 2021, 5.
- [5]. T. Ewertowski, M. Berlik, M. Sławińska, The Concept of Assessment of the Task Load of the Operator in the Aspect of Improvement of the Human-Technical-Environmental System on the Example of a Pilot, Report, *Poznan University of Technology*, 2020 (in Polish).
- [6]. L Salvan, S. T. Paul, A. Marois, Dry EEG-based mental workload prediction for aviation, in *Proceedings of the IEEE/AIAA 42nd Digital Avionics Systems Conference* (*DASC'23*), Barcelona, Spain, 2023, pp. 1-8.
- [7]. C. Tremmel, C. Herff, T. Sato, K. Rechowicz, Y. Yamani, D. J. Krusienski, Estimating cognitive workload in an interactive virtual reality environment using EEG, *Front. Hum. Neurosci.*, Vol. 13, 2019, 401.
- [8]. P. Zarjam, J. Epps, N. H. Lovell, Beyond subjective self-rating: EEG signal classification of cognitive workload, *IEEE Transactions on Autonomous Mental Development*, Vol. 7, 2015, pp. 301-310.

- [9]. S. Chikhi, S. Blanchet, N. Matton, EEG power spectral measures of cognitive workload: A meta analysis, *Psychophysiology*, Vol. 59, 2022 Jun, Issue 6, e14009.
- [10]. X. Cao, P. MacNaughton, L. R. Cadet, et al., Heart Rate Variability and Performance of Commercial Airline Pilots during Flight Simulations, *Harvard University*, Massachusetts, USA, 2019.
- [11]. L. Ji, L. Yi, H. Li, et al., Detection of pilots' psychological workload during turning phases using EEG characteristics, *Sensors (Basel)*, Vol. 24, Issue 16, 2024 Aug. 10, 5176.
- [12]. C. Liu, C. Zhang, L. Sun, et al., Detection of pilot's mental workload using a wireless EEG headset in airfield traffic pattern tasks, *Entropy*, Vol. 25, Issue 7, 2023, 1035.
- [13]. Y. Liu, Y. Gao, L. Yue, et al., A real-time detection of pilot workload using low-interference devices, *Appl. Sci.*, Vol. 14, Issue 15, 2024, 6521.
- [14]. M. Hamilton, The assessment of anxiety states by rating, Br. J. Med. Psychol., Vol. 32, 1959, pp. 50-55.
- [15]. K. Elnaggar, M. El-Gayar, M. Elmogy, Depression detection and diagnosis based on Electroencephalogram (EEG) analysis: a comprehensive review, *Diagnostics*, Vol. 15, Issue 2, Jan. 2025, 210.
- [16]. S. Yasin, A. Othmani, I. Raza, S. A. Hussain, Machine learning based approaches for clinical and non-clinical depression recognition and depression relapse prediction using audiovisual and EEG modalities: A comprehensive review, *Computers in Biology and Medicine*, Vol. 159, Jun. 2023, 106741.
- [17]. R. S. Aldahr, M. Alanazi, M. Ilyas, Evolving deep learning models for epilepsy diagnosis in data scarcity context: a survey, in *Proceedings of the 45th International Conference on Telecommunications and Signal Processing (TSP'22)*, Jul. 2022, pp. 66-73.
- [18] L. J. Bonnett, L. Kim, A. Johnson, J. W. Sander, N. Lawn, E. Beghi, M. Leone, A. G. Marson, Risk of seizure recurrence in people with single seizures and early epilepsy – Model development and external validation, *Seizure*, Vol. 94, Jan. 2022, pp. 26-32.
- [19] M. R. Islam, X. Zhao, Y. Miao, H. Sugano, T. Tanaka, Epileptic seizure focus detection from interictal electroencephalogram: A survey, *Cognitive Neurodynamics*, Vol. 17, Issue 1, Feb. 2023, pp. 1-23.
- [20]. C. R. Dhivyaa, K. Nithya, S. Anbukkarasi, Enhancing Parkinson's disease detection and diagnosis: a survey of integrative approaches across diverse modalities, *IEEE Access*, Vol. 12, 2024, pp. 158999-159024.
- [21]. A. M. Maitin, et al., Survey of machine learning techniques in the analysis of EEG signals for Parkinson's disease: a systematic review, *Applied Sciences*, Vol. 12, Issue 14, Jan. 2022, 6967.

- [22]. G. Greiner, Y. Zhang, Multi-modal EEG NEO-FFI with Trained Attention Layer (MENTAL) for mental disorder prediction, *Brain Informatics*, Vol. 11, Issue 1, Dec. 2024, 26.
- [23]. A. Tyagi, V. P. Singh, M. M. Gore, Towards artificial intelligence in mental health: A comprehensive survey on the detection of schizophrenia, *Multimedia Tools and Applications*, Vol. 82, Issue 13, May 2023, pp. 20343- 20405.
- [24]. S. Gomez-Quintana, A. O'Shea, A. Factor, E. Popovici, A. Temko, A method for AI assisted human interpretation of neonatal EEG, *Scientific Reports*, Vol. 12, Issue 1, Jun. 2022, 10932.
- [25]. M. Zhou, L. Zhou, M. Pan, X. Chen, An emotion recognition model based on long short-term memory networks and EEG signals and its application in parametric design, *Journal of Mechanics in Medicine and Biology*, Vol. 23, Issue 09, Nov. 2023, 2340096.
- [26]. T. Tian, L. Wang, M. Luo, W. Zhu, A novel psychotherapy effect detector of public art based on ResNet and EEG imaging, *Computational and Mathematical Methods in Medicine*, Apr. 2022, Vol. 2022, 4909294.
- [27]. F. Ma, Y. Yuan, Y. Xie, H. Ren, I. Liu, Y. He, F. Ren, F. R. Yu, S. Ni, Generative technology for human emotion recognition: A scoping review, *Information Fusion*, Vol. 115, Mar. 2025, 102753.
- [28]. Th. Mulder, Motor imagery and action observation: Cognitive tools for rehabilitation, *Journal of Neural Transmission*, Vol. 114, Issue 10, Oct. 2007, pp. 1265-1278.
- [29]. X. Wang, V. Liesaputra, Z. Liu, Y. Wang, Z. Huang, An in-depth survey on Deep Learning-based Motor Imagery Electroencephalogram (EEG) classification, *Artificial Intelligence in Medicine*, Vol. 147, Jan. 2024, 102738.
- [31]. C. M. Yilmaz, B. H. Yilmaz, Advancements in image feature-based classification of motor imagery EEG data: a comprehensive review, *Traitement du Signal*, Vol. 40, Issue 5, Oct. 2023.
- [32]. D. Das Chakladar, P. P. Roy, Cognitive workload estimation using physiological measures: A review, *Cognitive Neurodynamics*, Vol. 18, Issue 4, Aug. 2024, pp. 1445-1465.
- [33]. K. Kyriaki, D. Koukopoulos, C. A. Fidas, A comprehensive survey of EEG preprocessing methods for cognitive load assessment, *IEEE Access*, Vol. 12, 2024, pp. 23466-23489.
- [34]. Q. Sun, Y. Zhou, P. Gong, D. Zhang, Attention detection using EEG signals and machine learning: a review, *Machine Intelligence Research*, Jan. 2025.

(063)

Neurorehabilitation System Supported by Virtual Reality

<u>P. Śniatała</u>¹, S. Michalak^{2,3}, E. Kozielewska-Zwierska^{2,3}, A. Krawczyński³, K. Śniatała⁴ and S. Baliński¹

¹ Poznan University of Technology, Piotrowo 2, 60-965 Poznan, Poland
 ² Poznan University of Medical Sciences Przybyszewskiego 49, 60-355 Poznan, Poland
 ³ Institute of Neurological Disorders, Przybyszewskiego 49, 60-355 Poznan, Poland
 ⁴ Provincial Hospital in Poznan, ul. Juraszów 7/19, Poznań 60-479, Poland
 E-mail: pawel.sniatala@put.poznan.pl

Summary: Virtual reality has become an innovative method that finds applications in Health Care. One of the most popular applications of VR/AR in the health care area is Digital Therapeutics (DTx). The paper presents a rehabilitation system using virtual reality (VR) and augmented reality (AR). AR, implemented as a 'Smart Mirror', is used as an intelligent interface that, once a person is recognized, personalizes the subsequent dialogue with the user. Next, VR offers a dedicated, for the person set of rehabilitation exercises. The elaborated VR/AR rehabilitation system is currently used in everyday practice in Institute of Neurological Disorders at Poznan University of Medical Sciences. We have offered this approach to more than 100 post-stroke patients. To determine the tolerance of patients to the new rehabilitation technique, we conducted a survey and measurement of selected parameters of vital signs. The elaborate questionnaire included the presence of vertigo, nausea, diplopia, headache, chest pain, arrhythmia, anxiety, and sweating before and after VR training. We have also asked whether the patients are familiar with the use of computers / smartphones / games at home, and the educational level and profession were considered. Measurements of the tolerance and effectiveness (NHISS) show that the proposed solution supports the neurorehabilitation process well.

Keywords: Metaverse, Neurology, Rehabilitation, Augmented reality, Virtual reality.

1. Introduction

Healthcare is one of the most important contributors to the overall physical, social, and mental well-being of people around the world. Augmented and virtual reality in the healthcare market has been valued at more than \$2.5 billion in 2022 and is projected to register more than 21 % CAGR in the forest period [1]. Emerging new technologies are being used, when possible, in the healthcare field. Metaverse is one of these technologies, so it is being used in many areas of healthcare care [2, 3].

Neurological disorders represent one of the leading causes of disability and death globally. The burden of disability and mortality due to nervous system pathologies is steadily increasing, with recent data showing a 39 % increase in deaths related to neurological diseases in the past three decades. In 2021, 43 % of the global population, approximately 3.4 billion individuals, was reported to be affected by neurological conditions, a much higher figure than previously estimated [4]. These statistics underscore the urgent need for innovative solutions to address the growing global burden of neurological diseases.

One of the most popular applications of VR/AR in the Health Care area is Digital Therapeutics (DTx). It can be defined as evidence-based therapeutic interventions driven by software to prevent, manage, or treat a medical disorder or disease. In other words, DTx are patient-facing software applications that help patients treat, prevent or manage a disease and that have proven clinical benefit. DTx is expanding greatly, providing metaverse healthcare opportunities. This field offers great use of cognitive therapy, support groups, psychiatric assessments, and rehabilitation with the help of haptic sensors. Physical therapy is easy and responsive using AR and VR in the metaverse of healthcare. As presented in Fig. 1 neurology and psychiatry are the main areas of medicine where DTx is applicable.



Fig. 1. Neurology and psychiatry as the main areas of DTx applications.

Virtual reality has become an innovative method of rehabilitation over the past decade. By simulating everyday activities, stroke survivors can improve their self-care skills in a way that is usually not possible in a hospital setting. This article presents a virtual reality supported neurorehabilitation system. The innovative part of it is a patient communication environment using the 'Smart Mirror' developed and implemented by the authors. The system is designed to be used for patients and/or residents of a nursing home. In our case, it is dedicated to a group of patients who are undergoing rehabilitation after a stroke. The 'Smart Mirror' uses augmented reality to communicate with the patient and then the system can propose virtual reality exercises in the set of rehabilitation exercises. Thus, the system is an example of metaverse (AR/VR) application in the process of patient rehabilitation. The general idea of the proposed system is illustrated in Fig. 2.



Fig. 2. Illustration of the proposed system implementation.

2. Virtual Rehabilitation System

2.1. Medical Background

The rehabilitation process begins in the early stages of acute cerebral ischemia. Later, it continues during the patient's stay in the stroke unit. Disabilities caused by stroke vary, so it is necessary to adapt post-stroke rehabilitation to the needs of post-stroke patients and their daily activities at home. Most stroke survivors are discharged home. This raises the question of how to design rehabilitation training suitable for use at home. To initiate the adaptation process, some stroke centers (such as the University Hospital in Poznan, Poland) organize "model apartments" for initial patient training. Using various objects such as cabinets, sink, irons, spoons, cups, etc. as physical therapy tools in the "model apartment" supporting the patient's daily activities at home. As technology becomes more common and familiar, it can support rehabilitation in the hospital and/or home environment.

2.2. Virtual Rehabilitation System

Currently, the developed system is used in the rehabilitation of neurological patients (e.g., after strokes). As is well known, the rehabilitation process in the case of neurological diseases is a long process, and after a period of hospitalization, then, already at home, the patient must independently perform, often tedious, various exercises. Some of these exercises can be proposed to be performed in virtual reality. Performing these exercises in virtual reality helps make them more attractive and thus more effective (the patient is more likely to perform the exercises in a properly designed and attractive virtual reality). Of course, this is assuming that the motor requirements appropriate to the set of exercises are met.

The use of technology and virtual reality enables the dissemination of neurorehabilitation, enabling reaching such goals like:

- Physical and cognitive restoration;
- Enrichment of rehabilitation techniques;
- Increased effectiveness
- •Wider access to society;
- Teleneurorehabilitation.

Neurorehabilitation, supported with simple, everyday activity tools and advanced technology, opens possibilities that improve motor training and cognitive functions and enables the treatment of crippling symptoms like neglect syndrome.

2.3. Smart Mirror

The Smart Mirror (presented in Fig. 3) allows the patient to be identified through facial recognition, and then there is a personalized dialogue with the patient. Upon recognition of the patient, the system proposes to the patient a set of exercises dedicated to him (previously planned by a doctor and/or rehabilitant).

The proposal is not only personalized for the patient, but also takes into account additional information such as the time of exercise, current weather conditions (weather forecast), the patient's mood, etc. The substantive scope of the system software is consulted with neurologists and neurorehabilitation specialists and implemented in the system accordingly. After recognition of the patient, the system proposes to the patient a set of exercises dedicated to him (previously planned by a doctor and/or rehabilitant).

The system runs on a Raspberry Pi platform. However, the platform communicates with a more powerful computer that runs the main application. It is a web application implemented in the Django environment. Face recognition is implemented using the PyTorch and OpenCV libraries. The required data are stored in an SQL database (in our case MariaDB). The software structure of the system is shown in Fig. 4.



Fig. 3. Smart Mirror implementation.



Fig. 4. Software environments used in the application.

The facial recognition function in Smart Mirror is powered by algorithms based on artificial intelligence that can recognize faces and compare them with faces in the system database.

2.3. Virtual Exercises

The virtual rehabilitation system will propose the most appropriate and/or expected virtual interaction with the recognized person. The illustration of the system usage scenario is presented in Fig. 5.



Fig. 5. Example of the system usage scenario.

Those who plan the rehabilitation process can create a library of virtual activities dedicated to a particular patient. These can be games/activities that improve the patient's motor skills, or games that exercise mental activity. We have proposed to our patients simple games/exercises like for example a virtual fishing or origami. In fact, origami is a good example, since it requires hand-eye coordination, develops fine motor skills, and supports mental concentration, all of which stimulate the brain. The origami exercise is done in virtual reality. Of course, there are many other existing applications that can be used for rehabilitation purposes [5].

3. Clinical Rehabilitation Experience

The VR Rehabilitation System is currently used in everyday practice in the Stroke Unit of the University Hospital in Poznan. We have offered this approach to more than 100 patients after stroke. Disability caused by stroke vary and there is a need for personalized post-stroke rehabilitation. The success of post-stroke rehabilitation depends on many factors among them: patient pre-stroke activity, circulatory sufficiency, the severity and location of the stroke, and support from family and caregivers. Active participation of the stroke patient in the rehabilitation process is crucial to optimal recovery. Stroke patients are mainly older people who are only sometimes familiar with computer technology. In our pilot program with VR rehabilitation, we wanted to answer the following questions:

- What is the tolerance of this approach?
- What is its effectiveness?

To determine the tolerance of patients to the new rehabilitation technique, we conducted a survey and measurement of selected parameters of vital signs.

The elaborate questionnaire included the presence of vertigo, nausea, diplopia, headache, chest pain, arrhythmia, anxiety, and sweating before and after VR training. We have also asked whether the patients are familiar with the use of computers / smartphones / games at home, and the educational level and profession were considered.

Blood pressure, heart rate, and ECG are monitored before and after VR training. The evaluation of the patients was performed at baseline and after 7 days of training. The effectiveness of VR rehabilitation was measured using the National Institutes of Health Stroke Scale (NIHSS). The NIHSS is a 15-item neurological examination stroke scale used to assess the effect of acute cerebral infarction on levels of consciousness, language, neglect, loss of visual field, extraocular movement, motor strength, ataxia, dysarthria, and sensory loss. Ratings for each item are scored on a 3- to 5-point scale, with 0 as normal, and there is an allowance for untestable items. Scores range from 0 to 42, with higher scores indicating greater severity. Stroke severity may be stratified on the basis of NIHSS scores as follows [6]:

- Very Severe: 21–42;
- Severe: 16–20;
- Mild to Moderately Severe: 5–15;
- Mild: 1-4.

A trained observer rates the patent's ability to answer questions and perform activities, without coaching and without making assumptions about what the patient can do.

Based on the results of 102 patients involved in the program, we can conclude that there are no differences in questionnaire items before and after the VR neurorehabilitation. The same we can conclude based on the measured vital signs. Figs. 6-8, present the statistical distribution of Systolic Blood Pressure (SBP), Diastolic Blood Pleasure (DBP) and Heart Rate measurements before (SBP1, DBP1, HR1) and after (SBP2, SBP2, HR2) rehabilitation activities in the VR environment.

The results obtained show a very good response of patients to exercise using VR technology. Both the feedback received in the questionnaires and the measurements of blood pressure and heart rate did not reveal disturbing information regarding the patient's tolerance to this type of exercise.



Fig. 6. SBP before (1) and after (2) exercises.



Fig. 7. DBP before (1) and after (2) exercises.



Fig. 8. HR before (1) and after (2) exercises.

Lastly, the most importantly, the comparison of the NIHSS (as presented in Fig. 9) shows the effect of the proposed VR based rehabilitation. The NIHSS was improved from 4.707 to 2.155.



Fig. 9. The comparison of the NIHSS before (1) and after (2) the exercises.

We can conclude that the exercises carried out with patients using the described system produce positive results and tolerance is acceptable.

4. Conclusions

VR-based therapy can provide a positive learning experience, and be engaging and motivating exercises carried out with patients using the described system

produce positive results and the tolerance is acceptable. Further expansion of the palette of exercises available in virtual reality will reach homes of patients (e.g., elderly persons) and will allow telerehabilitation. Telerehabilitation trainings can be developed and supported by the implementation of human–computer interfaces (HCI).

Over the past decade, virtual reality has become a new way of rehabilitation from stroke and a unique method of treatment. By replicating actual activities, people recovering from stroke can perform self-care tasks in an environment that is usually impossible to recreate in the hospital environment. Virtual reality is increasingly being used in this context and its potential medical applications are still not fully understood. The profound impact on stroke survivors is obvious, as they use VR technology to recreate important daily activities, promote new neural connections, and improve their self-confidence. VR-based stroke games are known to increase the attendance of the patients, boost their morale, and provide them with selfconfidence to carry on with their lives. Virtual reality exercises for stroke are known to speed up the recovery process, provide muscle strength, and also bring balance to the body. VR-based journeys are known to provide calm and relaxation to patients' stressful minds. As more and more survivors use this technology to retrain their limbs, the future of virtual reality in stroke recovery looks promising. VR-based therapy can provide a positive learning experience and be engaging and motivating.

The review [7] shows that telerehabilitation in stroke patients is superior or similar to conventional rehabilitation in clinical outcomes and is used as a complementary therapy or as alternative treatments. More importantly, TR provides access to rehabilitation services for a large number of patients with immobility, living in remote areas, and during the COVID-19 pandemic or similar events. As is usually the case with new technologies, it is important to strike a balance and detect the good points of new developments. If they can contribute to saving health, they should be used, keeping in mind the principle of doing no harm.

References

- A. M. Al-Ghaili, H. Kasim, N. M. Al-Hada, et al., A review of metaverse's definitions, architecture, applications, challenges, issues, solutions, and future trends, *IEEE Access*, Vol. 10, 2022, pp. 125835-125866.
- [2]. R. Chengoden, N. Victor, T. Huynh-The, et al., Metaverse for Healthcare: A survey on potential applications, challenges and future directions, *IEEE Access*, Vol. 11, 2023, pp. 12764-12794.
- [3]. B. Scheffler, F. Schimböck, A. Schöler, K. Rösner, J. Spallek, C Kopkow, Tailored guideline implementation in STrokE Rehabilitation (GLISTER) in Germany. Protocol of a mixed methods study using the behavior change wheel and the theoretical domains framework, *Front. Neurol.*, Vol. 13, 2022 Jul 27, 828521.
- [4]. J. D. Steinmetz, K. M. Seeher, N. Schiess, et al., Global, regional, and national burden of disorders affecting the nervous system, 1990–2021: a systematic analysis for the global burden of disease study 2021, *The Lancet Neurology*, Vol. 23, Issue 4, 2024, pp. 344-381.
- [5]. K. E. Laver, B. Lange, S. George, J. E. Deutsch, G. Saposnik, M. Crotty, Virtual reality for stroke rehabilitation, *Cochrane Database of Systematic Reviews*. Vol. 11, 2017, CD008349.
- [6]. T. Brott, H. P. Adams Jr., C. P. Olinger, J. R. Marler, et al., Measurements of acute cerebral infarction: a clinical examination scale, *Stroke*, Vol. 20, 1989, Issue 7, pp. 864-870.
- [7]. V. A. Nikolaev, A. A. Nikolaev, Recent trends in telerehabilitation of stroke patients: A narrative review, *NeuroRehabilitation*, Vol. 51, 2022, Issue 1, pp. 1-22.

(064)

Radioactive Tabular Datasets to Detect Unauthorized Machine Learning

Mehdi Ben Ghali^{1,2,3}, Gouenou Coatrieux1,² and Reda Bellafqira^{1,2}

¹Inserm UMR 1101 LaTIM, Brest, France ² IMT Atlantique, 665 Technopôle Av., Brest, France ³ Inserm Grand Ouest, Nantes, France Tel.: + 33 0612468754 E-mail: {ben-ghali, gouenou.coatrieux, reda.bellafiqra}@imt-atlantique.fr

Summary: Being able to prove a dataset was used to train a particular deep learning model is a real need that can be used to demonstrate dataset unauthorized use or reuse. It is also a technical challenge. Recently, radioactive data techniques which modify data so that it leaves a trace in any model trained on it have been proposed to solve this challenge for image datasets. But they have yet to be extended to other domains. In this paper, we introduce R-TAB, the first technique implementing the concept of radioactive data for tabular datasets. R-TAB is a radioactive-based approach that modifies selected database attributes under correlation constraints to leave a retrievable trace in any model trained on these data. Experiments conducted on several datasets and models demonstrate that our solution is robust in terms of radioactivity detection while maintaining model training performance. Finally, we provide an analysis of constraints and criteria such techniques for tabular datasets have to consider going forward.

Keywords: Deep learning, Radioactive data, Databases, Tabular data, Ownership protection.

1. Introduction

As Deep Learning models are trained on ever-increasing amounts of data, the issue of unauthorized collection and use or re-use of datasets has become a serious challenge. While legal frameworks and laws protect against such unconsented exploitation of data [1], in practice, it remains very hard to detect that a dataset has been used for the training of a particular model. If some solutions based on a posteriori membership inference techniques have been proposed [2], they appear inaccurate and demand heavy resources and knowledge of the model. More recently, a few techniques based on the concept of radioactive data [3] have been introduced. They rely on the injection of data isotopes [4] which are slightly modified versions of the original data, with as an objective to leave an identifiable trace in the model trained on these data. The trace then serves as proof that the model was trained on a given dataset. To the best of our knowledge, radioactive data techniques focus on image datasets. No proposal exists yet for tabular data, which remains omnipresent in many ownership-sensitive domains like finance and healthcare. In this paper, we introduce the first technique to generate radioactive tabular data while underlying the main constraints to satisfy. The rest of this paper is organized as follows. Section 2 comes back on image radioactive techniques. Section 3 details our scheme. Section 4 provides some experimental results demonstrating our solution detects radioactive models trained on some tabular data while preserving the accuracy of the model on its main task. Section 5 provides points of discussion. Section 6 concludes this paper.

2. Related Works

Sableyrolles *et al.* [5] were the first to introduce the concept of radioactive data for image classification. By modifying some images in a dataset, they align the classifier layer of the model with a secret vector. For verification, they compute the cosine distance of the classifier of a model to a secret vector and, if it is high enough, the proof is given that the model was trained on the isotope samples. Solutions proposed subsequently embrace the same principles.

Existing solutions from literature can fit into one of four categories depending on two criteria: the way radioactive isotopes of samples are generated and the inspection technique of the trained model, both being linked. We thus suggest the following classification:

- White-box vs. Black-box radioactive methods To detect the trace left by the radioactive dataset, some solutions follow a white-box approach by inspecting the model parameters, like [5] which examines the direction of the classifier layer. Other solutions, referred to as black-box schemes, such as [4] or [6] only need access to the outputs of the model to decide if this one has been trained on a radioactive image dataset.
- Fixed vs. Guided radioactive methods To generate radioactive isotopes, these solutions rely on two classes of techniques: i) the insertion of a predefined noise or mark in the original samples; ii) the modification of said samples under some objective function or constraint. We propose to refer to the former category as "fixed" isotope creation, the second as "guided" isotope creation. An example of a fixed black-box radioactive scheme is data isotopes [4]. Their idea is to add a mark to images pertaining to one target class so

as to add artificial features to this class for insertion. The radioactive model will associate these features to this class. At inspection, the trained model is fed with images from another class both with and without the mark. If the model was trained on the isotopes, an increase towards the radioactive class is observed in the output probabilities for these samples. To mark images, they superpose a predetermined image selected from outside the dataset over them. This is a fixed modification. For detection, they only need access to the probabilities output by the model. This is a black-box condition.

Table 1 provides the classification of the different radioactive image data methods we found in literature. We note that most solutions are guided black-box methods.

Table 1. Classification of radioactive image data techniques in literature. Note that the authors in [7] present both a black-box and white-box version of their scheme.

	White-box	Black-box
Guided	Sablayrolles et al. [5] Anti-Neuron* [7]	Metapoison [8] Anti-Neuron* [7] Untargeted Backdoor [9] Data Taggants [6]
Fixed	Catch Me If You Can [10]	Data Isotopes [4]

3. Proposed Method

The radioactive tabular dataset solution we propose, R-TAB, is a guided black-box scheme in the context of classification tasks. More specifically, it belongs to a subclass of black-box methods the radioactivity process of which consists in adding some characteristics or features, called "spurious" features, to samples from one dataset class, the radioactive class. In images, this corresponds to inserting content in unused regions of the image [4] or adding some noise [10]. Any model trained with the radioactive samples will learn to associate these features to the radioactive class. The idea is that if samples from another class to which these artificial characteristics have been added are presented to the radioactive model, a bias in class probabilities towards the radioactive class is expected. When detected, this bias proves that the model was trained on the radioactive dataset.

To detail how to adapt this concept to tabular data, let us consider a relational database constituted of a single table *T* of *N* tuples and *K* attributes: $T = \{t_u.A_v\}_{u=1..N,v=1..K}$, where each tuple t_u has a class label $t_u.C$. One of the main constraints to consider is that, unlike images, which can contain up to millions of pixels, a tuple is usually limited to a few attributes. It is therefore difficult to imagine being able to add noise to a single tuple. Thus, instead of adding a specific mark to each sample, we suggest inserting the spurious features in the statistical properties of one secretly selected class C_1 , more specifically in the joint distribution of its most relevant attributes. The idea then is to verify that the radioactive model, if inferred with the samples of another class C_j modified similarly as previously, provide higher C_1 probabilities for these samples. We detail these marking and detection processes below.

3.1. Radioactive Dataset Creation

This process works as follows:

- 1. Secretly select the radioactive class C_1 ;
- 2. Compute the pairwise squared Pearson correlation coefficient r [11] of each two attributes from A_1 to A_K , across C_1 samples;
- 3. Select the two most correlated attributes A_z and A_w ;
- 4. Transform C_1 samples into isotopes C_1^* by modifying A_z values so as to slightly de-correlate it from A_w while keeping its distribution change minimal. This can be achieved using a common gradient descent-based optimization process with the following loss function:
- 5. $\mathcal{L}(A_z, A_w) = r^2(A_z, A_w) + \lambda ||\mu_z -\mu_z^0||^2 + \alpha ||\sigma_w \sigma_w^0||^2;$

where μ and σ denote the attribute mean and standard deviation, respectively, and λ and α are weights to be tuned. The first term describes the correlation to reduce while the two others act as regularization to prevent the excessive distortion of the distribution of A_z . Once the radioactive dataset is created, it can be shared.

3.2. Model Inspection

To determine if a model *M* was trained using radioactive data, one just has to follow these steps:

- 1. Select tuples from a class C_2 different from C_1 ;
- 2. Make a radioactive version C_2^* of these tuples by applying the same process as above, on the same attributes A_z and A_w ;
- 3. Compute class probabilities by passing both versions of the samples to the model;
- 4. To decide whether a bias towards C_1 appears in the classification probabilities of radioactive tuples C_2^* , we perform a Student's one-sided paired samples t-test [12]. It outputs a p-value for the null hypothesis H_0 :"*The average class probability for* C_1 *is equal or less in the radioactive group*". If below a certain threshold, p-value confirms the model was trained on the radioactive dataset, highlighting a statistically significant increase in the average of class probabilities for C_1 for the radioactive samples.

4. Experimental Results

4.1. Datasets, Classification Task and Evaluation Criteria

For this version of our paper, we provide results on three references open-source classification datasets:

UCI Forest Cover type dataset [13], UCI EEG Eye State dataset [14] and the Turing Institue vehicle recognition dataset [15]. We will refer to them as Covertype, EEG and Vehicle. We provide descriptors of the datasets below:

- Covertype: Consists of cartographic data for several 30×30 meter forest cells and the corresponding cover (tree) type. The target task is to predict the type of tree present in a given cell from its geographic data;
- EEG: Tracks the values measured during a 117 second continuous electroencephalogram and the state of the patient's eye detected via a camera. The target task is to predict whether a patient's eye is closed or open given electroencephalogram measures;
- Vehicle: Contains the numerical descriptors of 2D silhouettes of four different vehicles under different angles. The silhouettes were captured by a camera and processed to extract numerical features. The target task is to classify a silhouette into the corresponding vehicle given its numerical descriptors. Please note that for simplicity purposes we worked with a simplified version of the dataset [16], where the task is only to determine if the given silhouette is that of a car or not.

Table 2 contains the descriptors of the three datasets.

Dataset	CoverType	EEG	Vehicle
No. Samples	581,012	14,980	98,528
No. Features	54	14	100
No. Classes	7	2	2
Feature Type	Numeric, Categorical	Numeric	Numeric

Table 2. Description of datasets used for our experiences.

As a model, we use the ResNet-inspired architecture for tabular data classification introduced in [17]. It works similarly to the ResNet architecture for image classification [18], with the difference that convolutional blocks are replaced with linear layers. They demonstrate that it serves as a good baseline architecture for a wide variety of tabular data classification tasks [17]. We adapt the size and number of blocks in the model for each dataset to what achieves good baseline scores in the absence of radioactivity (see Section 4.2). As a second model, we also use a simple multi-layer perceptron, made of three 192-parameter layers.

Two metrics were considered to evaluate the performance of our method:

- *Main classification task performance* (ACC) – As radioactivity should not harm the dataset usability, we track the accuracy (ACC Radioactive) of models trained on the radioactive *vs.* non-radioactive dataset on an unseen test set.

It should not decrease compared to the one of the model trained on non-radioactive data (ACC Base).

- Radioactive model detection (p-value) – The p-value corresponding to the null hypothesis H_0 (see Section 3.2) is used as a detection performance score. It should be small. We decided on a threshold of 0.1, which corresponds to a 90 % probability of the model being radioactive.

4.2. Experimental Setup

To obtain a baseline performance, we split each dataset into 60 % training data and 20 % validation and test data. Then, we manually perform hyperparameter search to settle on the best performing models and parameters. For Covertype, EEG and Vehicle respectfully, we settle on a (2,192,2), (3,192,2) and (3,1024,2) ResNet where the parentheses represent the number of hidden blocks, the first block's dimension and the following blocks dimension multiplier.

For the covertype dataset, we used "Spruce/Fir" as radioactive class (C_1) and "Lodgepole Pine" as inspection class (C_2). We created isotopes by modifying the "hillshade_9am" (A_z) attribute with regards to the "hillshade_3pm" attribute (A_w), performing 120 epochs and setting $\lambda = \alpha = 0.4$. For both EEG and Vehicle, we used the "0" class as a radioactive class and the "1" class for inspection. EEG isotopes were created by modifying the attribute "FC5" with regards to "O1" over 1000 epochs with a 2.0 learning rate and $\lambda = \alpha = 0.95$. Vehicle isotopes were created by modifying the attribute "X48" with regards to "X49" over 500 epochs with a 3.0 learning rate and $\lambda = \alpha = 0.95$.

4.3. Experimental Results

Table 3 reports the results we obtained in terms of baseline and radioactive models' accuracies and of model radioactivity detection. Notice that the p-value is to" X given in average over 3 trials like in [4], in order to offset the effects of random batching. It can be seen that, on different data sets of different content, classification purposes and sizes, our method verifies the two criteria. Accuracy on the test sets is preserved, while obtaining p-values under our fixed threshold in all experiments, *i.e.*, we detect that the model was trained on radioactive data every time.

Table 3. Experimental results for the cover type dataset.

Dataset	ACC Base	ACC Radioactive	p-value
CoverType	0.938	0.943	5.4 * 10-2
EEG	0.853	0.852	2.5 * 10-6
Vehicle	0.862	0.848	5.9 * 10 ⁻²

Since our method is model-agnostic. i.e., it does not make prior hypotheses on the type of model that will be trained on the data unlike methods such as [5] or [6]. We also tested it on a linear multi-layer perceptron (see Section 4.1). But, due to time constraints, we were only able to do so for the Vehicle datasets. Table 4 provides the performance results we obtained using the same radioactivity parameters as previously. It can be seen that our radioactive isotopes are still effective when training on another class of models than ResNet. It is reasonable to assume the architecture independence of our scheme.

 Table 4. Experimental results for the Vehicle dataset with an MLP architecture.

Dataset	ACC (base)	ACC (radioactive)	p-value
Vehicle	0.867	0.871	9.5 * 10 ⁻²¹

Since real-life applications of datasets go beyond just the training of machine learning models (e.g., statistical studies and data visualization). We also study the impact that our approach can have on different statistical properties of these data. We report in Table 5 the standard deviation and mean of the distribution of the modified attributes in each dataset. As shown, these properties are not heavily impacted by our method. In detail, the insertion of isotopes introduces a negligible distortion for both the CoverType and EEG datasets. For Vehicle, this shift is less negligible. But as illustrated in Fig. 1 which provides a point cloud visualization of the modified samples before and after applying radioactivity, the distribution of the radioactive feature values is still more or less preserved. From this standpoint, one can assume that our method should not harm the usability of altered datasets.

 Table 5. The impact of radioactivity on the statistics of the modified attribute for each dataset.

Deterret	Μ	ean	Standard deviation		
Dataset	Base	Isotope	Base	Isotope	
CoverType	21.998	21.999	24.820	24.821	
EEG	4200.391	4200.233	7026.078	7026.082	
Vehicle	-1.532	-2.105	1.108	1.276	

Finally, the other criterion to account for in real-life applicability is computational overhead. In machine learning experiments, training time scales with the size of data and models. More clearly, training epochs are longer when there are more samples, more features, and when the trained model has more trainable parameters. It is thus important that the injection cost of isotopes is not significant compared to the training of the model. Table 6 reports the different times taken by both the injection and classification tasks in our experiments. One can see that for all three datasets the radioactive modification time is negligible compared to model training, representing only percentiles of the duration of a single training epoch.



Fig. 1. Scatter plot of the "X48" attribute values of the radioactive class from the Vehicle dataset before and after modification (Each point corresponds to a single sample from the radioactive class. X-axis shows the attribute value. Y-axis is an arbitrary index for each sample).

Table 6. Comparison of computation time necessary for the different steps of the training of a radioactive model. Training is the time needed for one training epoch. Isotope is the time needed for the entire radioactivity process. Ratio is the fraction of training time it represents.

Dataset	Training	Isotope	Ratio
CoverType	79.62 s	0.84 s	1.05 %
EEG	49.38 s	2.67 s	5.4 %
Vehicle	17.617 s	0.17 s	0.96 %

5. Discussion

While the above experiments prove the effectiveness of our method, and that the concept of radioactive data can be applied to tabular datasets, several additional points have to be discussed for a more general application.

First, it is important to minimize data distortion when injecting isotopes for two reasons: the perceptibility of the radioactivity, and the usability of the dataset. Up to now, all radioactive image data methods in literature do not satisfy the former criterion, as modifications to images are easy to spot by a human observer and stand out from natural samples. Tabular data suffer less from this problem, because it is more difficult for a human being to interpret individual samples characterized by a few dozen attributes. Nevertheless, it is highly probable that the majority of users will carry out simple data statistics and analyses. It is therefore important that the radioactive data do not appear as anomalies at the statistical level. Regarding dataset usability and our scheme which creates radioactive isotopes by adding noise through an optimization process, there exists an implicit lower bound for the loss value which corresponds to a level of noise beyond which the data usability degrades. We enforce the preservation of the distribution of data; through mean and standard deviation; as a usability-preserving criterion, but it may not be sufficient in more complex analyses.

The robustness of radioactive data in the face of modifications, even if this was beyond the scope of our article, is another property to be taken into account. One should at least consider non-malicious modifications, that is to say data processing that does not aim to remove the radioactive isotopes, but which could nevertheless have an impact on them. More clearly, it is standard in machine learning to preprocess data before using it for training, by normalizing it or removing outliers for example. In the context of radioactive images, [5] takes into account common preprocessing operations on images for example.

In this work, we propose the first radioactive protection for datasets. It is complementary to other dataset protection tools for ownership verification techniques such as watermarking and membership inference.

Watermarking is mainly devoted to the ownership protection of dataset or to the fight against information leaks. Many solutions have been proposed to watermark tabular data [19-21]. Its basic principle consists in inserting a watermark (equivalently a message or a proof of ownership or the recipient's identifier) in the dataset by modifying its attributes' values or by injecting false attributes. However, unlike radioactive isotopes, the watermark does not inherently transfer to models trained on watermarked data. However, both mechanisms introduce noise in the dataset, in general.

On their side, membership inference methods aim to determine whether a specific sample was part of the training set. They usually assume that a model shows stronger responses in terms of higher-class probability, higher activation in neurons, and so on, for samples seen during its training phase. As stated, membership inference and radioactive data both rely on the fact model overfit some training data characteristics. They are however completely different in their functioning, namely in the level of knowledge and access the user has over the dataset to be tracked [3]. Radioactive data is a dataset-level technique where the dataset owner injects the spurious features to be memorized by the model; features unlikely to be naturally present in samples. Membership inference tools do not use such features. Moreover, even though a membership inference can identify samples from the training data, it does not mean a specific dataset has been used as several datasets can share several samples. That is the case of medical records for example. As it is not uncommon for a patient to have visited several hospitals, which means that their data appears in several datasets.

6. Conclusion

In this work, we introduce R-TAB, the first radioactive data approach for tabular datasets to detect their unauthorized use for training machine learning models. It takes into account the specificities of tabular datasets to extend similar schemes proposed for images. As another originality, it creates isotopes by introducing spurious features in some secret attributes while preserving attributes' correlation. We tested the method as a proof of concept on three real datasets for classification applications and with two kinds of models: ResNet and MLP. As demonstrated, our method preserves the model accuracy while allowing good detection. We also discuss some of the constraints to satisfy when creating radioactive datasets. Future work will focus on how to better preserve dataset usability and to the generalization of our approach to other kinds of machine learning algorithms.

Acknowledgements

This work was partly supported by a French government grant managed by the Agence Nationale de la Recherche under the France 2030 program, reference ANR-22-PESN-0006, and by a Brittany Council and European FEDER through the industrial chair CYBAILE.

References

- [1]. A. Mantelero, The EU Proposal for a General Data Protection Regulation and the Roots of the 'Right to Be Forgotten', *Rochester*, 2013.
- [2]. H. Hu, Z. Salčić, G. Dobbie, J. Chen, L. Sun, X. Zhang, Membership inference via backdooring, in Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, Vienna, 2022.
- [3]. L. Du, X. Zhou, M. Chen, et al., SoK: dataset copyright auditing in machine learning systems, *arXiv preprint*, 2024, arXiv:2410.16618.
- [4]. E. Wenger, X. Li, B. Y. Zhao, V. Shmatikov, Data isotopes for data provenance in DNNs, *arXiv preprint*, 2023, arXiv:2208.13893.
- [5]. A. Sablayrolles, M. Douze, C. Schmid, H. Jegou, Radioactive data: tracing through training, in *Proceedings of the 37th International Conference on Machine Learning (ICML'20)*, 2020, pp. 8326-8335.
- [6]. W. Bouaziz, E.-M. El-Mhamdi, N. Usunier, Data taggants: dataset ownership verification via harmless targeted data poisoning, *arXiv preprint*, 2024, arXiv:2410.09101.
- [7]. Z. Zou, B. Gong, L. Wang, Anti-neuron watermarking: protecting personal data against unauthorized neural networks, https://www.ecva.net/papers/eccv_2022/ papers_ECCV/papers/136730449.pdf&ved=2ahUKE wj9pPCx35uMAxXU_rsIHe4nHWEQFnoECBwQAQ &usg=AOvVaw1snvdYkDkY_RYNavjHN8Az
- [8]. W. R. Huang, J. Geiping, L. Fowl, et al., MetaPoison: practical general-purpose clean-label data poisoning, *arXiv preprint*, 2020, arXiv:2004.00225.

- [9]. Y. Li, Y. Bai, Y. Jiang, Y. Yang, S.-T. Xia, et al., Untargeted backdoor watermark: towards harmless and stealthy dataset copyright protection, *arXiv preprint*, 2022, arXiv:2210.00875.
- [10]. Z. Chen, K. Pattabiraman, Catch me if you can: detecting unauthorized data use in deep learning models, arXiv preprint, 2024, arXiv:2409.06280.
- [11]. H. E. Soper, A. W. Young, B. M. Cave, A. Lee, K. Pearson, On the distribution of the correlation coefficient in small samples. Appendix II to the papers of "student" and R. A. Fisher. A cooperative study, *Biometrika*, Vol. 11, Issue 4, May 1917, pp. 328-413.
- [12]. Student, The probable error of a mean, *Biometrika*, Vol. 6, 1908, pp. 1-25.
- [13]. J. Blackard, Covertype, UCI Machine Learning Repository, 1998.
- [14]. O. Roesler, EEG Eye State, UCI Machine Learning Repository, 2013.
- [15]. P. Mowforth, B. Shepherd, Statlog (Vehicle Silhouettes), *UCI Machine Learning Repository*.
- [16]. J. Siebert, Vehicle Recognition Using Rule Based Methods, Research Memorandum TIRM-87-018, *Turing Institute*, 1987.

- [17]. Y. Gorishniy, I. Rubachev, V. Khrulkov, A. Babenko, Revisiting deep learning models for tabular data, *arXiv* preprint, 2021, arXiv:2106.11959.
- [18]. K. He, X. Zhang, S. Ren, J. Sun, M. Research, Deep residual learning for image recognition, *arXiv preprint*, 2015, arXiv:1512.03385.
- [19]. J. Franco-Contreras, et al., Adapted quantization index modulation for database watermarking, *Lecture Notes in Computer Science*, Vol. 9023, 2015, pp. 120-134.
- [20]. J. Franco-Contreras, G. Coatrieux, F. Cuppens, N. Cuppens-Boulahia, C. Roux, Robust lossless watermarking of relational databases based on circular histogram modulation, *IEEE Transactions on Information Forensics and Security*, Vol. 9, Issue 3, March 2014, pp. 397-410.
- [21]. D. Niyitegeka, G. Coatrieux, R. Bellafqira, E. Genin, J. Franco-Contreras, Dynamic watermarking-based integrity protection of homomorphically encrypted databases-application to outsourced genetic data, in *Proceedings of the International Workshop on Digital Watermarking*, 2018, pp. 151-166.

(065)

MineralBLIP: Advancing Mineral Classification with Vision Language Pre-training Model

<u>Khalid Alharthi</u>¹, Ghadi Alkhushail¹, Sharifah Malhan¹, Batol Alsalkhadi¹, Hatun Alqarni¹, Kholoud Alharthi¹, Reem Almarhabi¹, Raghad Alharthi¹, Ali Alshahrani¹, Muhammad Zaka Emad² and Dhafer Alshehri²

 ¹ Department of Computer Science, College of Computing, University of Bisha, Bisha 61922, P.O. Box 551, Saudi Arabia
 ² Department of Petroleum Engineering, King Fahad University of Petroleum and Minerals, Dhahran 31216, Saudi Arabia Tel.: +966 53 644 4911 E-mail: <u>kharthi@ub.edu.sa</u>

Summary: Accurate mineral identification and classification is crucial for effective exploration, resource management, and industrial applications in resource-rich regions. Recent advances in artificial intelligence (AI) have enabled automated mineral classification, with deep learning approaches such as Convolutional Neural Networks (CNNs) widely adopted. However, conventional CNN-based models often fail to capture fine-grained features, limiting their performance in complex scenarios. Motivated by breakthroughs in image captioning via Bootstrapped Language-Image Pre-training, we propose a novel Vision-Language Pre-training (VLP) approach, MineralBLIP, to improve mineral classification accuracy. MineralBLIP employs a multimodal framework that integrates computer vision and natural language processing techniques. Experimental evaluations on two mineral image datasets demonstrate that MineralBLIP achieves an average F1-score of 84 %, markedly surpassing the CNN model's 75 %. These results underscore the promise of vision-language models in advancing mineral classification research and the role of advanced AI in mineral identification and classification research leading to sustainable mine development.

Keywords: Minerals image classification, CNN, Vision language pre-training, BLIP.

1. Introduction

Mineral classification is vital for successful mining, resource management, and industrial applications, as accurate identification informs resource estimation, environmental assessment, and mineral processing. Traditional methods rely on physical inspection and manual analysis, which are time-consuming and prone to human error [1]. Consequently, the field has increasingly turned to automated approaches using machine learning (ML) and deep learning (DL) techniques to extract meaningful features from mineral images.

Convolutional Neural Networks (CNNs) have been widely used for mineral classification due to their ability to learn spatial hierarchies in images and capture patterns in visual data [2, 3]. Several studies have demonstrated the success of CNNs in this domain [4, 5]. However, CNNs struggle with capturing long-range dependencies within an image as they rely on localized receptive fields, limiting their ability to recognize complex, context-dependent relationships in mineral images [6]. In addition, CNNs require large amounts of labeled data for effective training, which may not always be available in very specialized domains such as mineral and rock classification.

Recent advances in vision-language models (VLMs) have emerged as promising alternatives to traditional CNNs for image classification, particularly in complex tasks such as mineral image analysis. Unlike CNNs, VLMs integrate both visual and textual

modalities to enrich feature representation. Vision Transformer (ViT) architectures, for example, leverage self-attention mechanisms to capture global dependencies across an image, allowing them to learn relationships among distant pixels [7]. Building on this, models such as Bootstrapped Language-Image Pre-training (BLIP) incorporate a captioning and filtering process that extracts subtle semantic details even from limited datasets [8]. This multimodal approach effectively bridges the gaps left by conventional CNNs, offering improved transfer learning and fine-tuning capabilities for domainspecific tasks such as mineral classification [9-11].

In this paper, we harness the strengths of vision-language pre-training (VLP) by proposing MineralBLIP, a model that adapts the BLIP framework for mineral image classification. By combining the global feature extraction capabilities of Vision Transformers (ViTs) with BLIP's unique captioning and filtering process, MineralBLIP overcomes the limitations of CNN-based methods. Our experimental evaluation on two mineral image datasets demonstrates that MineralBLIP achieves an average F1-score of 84 %, substantially surpassing the corresponding CNN baselines of 75 %. Our analysis confirms that MineralBLIP not only achieves remarkable improvements in classification performance but also demonstrates significant computational efficiency over traditional CNN-based approaches. These results demonstrate the advantages of vision-language pre-training for domain-specific

tasks, particularly in data-scarce scenarios with subtle visual differences. MineralBLIP's improved performance promises benefits including optimized mining, streamlined mineral analysis, and cost reductions across industries.

The remainder of this paper is organized as follows: Section 2 reviews related work, Section 3 details the proposed methodology, Section 4 presents experimental results, and Section 5 summarizes our work.

2. Related Work

Deep learning has revolutionized image classification, in particular with convolutional neural networks (CNNs) and transformer-based architectures such as Vision Transformers (ViTs). CNNs have traditionally dominated this field due to their advanced feature identification capabilities for composite image data [12]. However, as classification tasks require modeling of long-range dependencies and global context relationships, ViTs have emerged as a powerful alternative. While CNNs remain widely used in important domains such as medical imaging and object recognition, ViTs excel with capturing complex spatial relationships, making them very effective for a range of applications [13].

CNNs are extensively applied in mineral classification, particularly in hyperspectral image analysis for mineral identification. Researchers have conducted studies on optimizing CNN models to enhance accuracy of classification. In this context, Attallah et al. [14] fine-tuned a 3D-2D CNN to enhance mineral classification, achieving improved feature extraction in hyperspectral mineral imaging. Brempong et al. [15] introduced MiNet, a lightweight CNN designed for real-time mineral recognition with reference to mining applications. Additionally, Cifuentes et al. [16] incorporated short-wave infrared (SWIR) hyperspectral imaging with CNNs, refining classification performance.

Despite CNNs' effectiveness, it faces challenges in differentiating minerals with similar spectral characteristics because of their localized feature extraction focus. This limitation affects their ability to model global contextual relationships, making it difficult to classify minerals with subtle texture variations [17].

To overcome these limitations of CNNs, ViTs offer self-attention mechanisms to capture both local and global dependencies, leading to an improved classification performance. Liu et al. [18] explored hybrid approaches integrating both ViTs with CNNs, significantly improving their capability to differentiate minerals with similar textures. He et al. [19] showed how effectively ViTs capture long-range dependencies in mineral textures, producing superior multi-label classification results. Liu et al. [20] studied ViTs ' capability to analyze complex spatial relationships, improving geological data interpretation. Beyond mineral classification, ViTs have been applied in broader fields such as geology and remote sensing. Koeshidayatullah et al. [21] introduced FaciesViT, a vision transformer model designed for lithofacies prediction, demonstrating its effectiveness in many geological applications.

Multimodal learning models, such as BLIP, have gained attraction for improving accuracy of image captioning by integrating textual and visual data. It has already been successfully applied for different domains (e.g., medical imaging classification [22]). Also, Xiao et al. [23] introduced a BLIP-2-based model designed for processing point cloud data, exhibiting its effectiveness in object recognition tasks. In addition, Tao et al. [24] studied BLIP-based image enhancement techniques, showing their potential in refining image classification in remote sensing. Also, as shown in [8], Li et al. demonstrated that ViTs, when combined with BLIP, could enhance feature extraction in hyperspectral image understanding. The integration of textual descriptions with visual data through models such as BLIP offers a promising direction for image captioning, enabling more precise and context-aware results.

Finally, Nguyen et al. [25] proposed hybrid approaches using Vision Transformers (ViTs) with CNNs to further enhance mineral classification performance, while Liu et al. [26] discussed forthcoming developments with respect to applying these hybrid approaches for complex mineralogical data modeling and analysis.

Differently from existing work, this study aims to show that our solution, MineralBLIP, exhibits significant potential in the field of mineral classification, as it is capable of effectively processing intricate image details and understanding the relationships between them.

3. Methodology

Accurate mineral classification is essential in fields such as geology, mining, and material science. While CNN-based models have shown effectiveness in general visual tasks, they face challenges in mineral classification due to high visual similarity among mineral types and the limited availability of labeled datasets. These limitations often lead to reduced classification accuracy and poor generalization.

To overcome these challenges, we introduce MineralBLIP, a BLIP-based model that integrates both visual and textual information to enhance classification performance. Unlike unimodal CNN models, which rely solely on visual features, visionlanguage models such as BLIP leverage both image and textual data, enabling improved interpretability, robustness, and generalization, even with limited datasets.

BLIP (Bootstrapped Language-Image Pretraining) is a state-of-the-art vision-language model that aligns visual and textual representations through a joint training process. It features a unique captioning and

filtering mechanism, where it generates descriptive captions for images and filters out noisy or irrelevant data. This mechanism significantly improves the model's ability to understand and classify images with fine-grained details, making it well-suited for tasks such as mineral classification.

In this study, we adapt the BLIP model to the task of mineral image classification by fine-tuning its vision- language pipeline. Our methodology follows a structured, multi-stage approach to adapt and fine-tune the BLIP model for classifying minerals based on their raw images. The process consists of three main phases:

3.1. Dataset Preparation Phase

Two datasets were used to evaluate the performance of MineralBLIP:

• Dataset 1 (from Mindat.org) contains 6172 images from four mineral classes: beryl, copper, malachite, and wulfenite [27];

• Dataset 2 (from Roboflow Universe) consists of 2310 images from five mineral classes: malachite, chrysocolla, quartz, pyrite, and muscovite [28].

Fig. 1 provides representative samples, labeled with both their names and chemical formulas.

- Both datasets underwent preprocessing as follows:
- The datasets were filtered to focus on the specified classes, ensuring a manageable and well-defined classification task;
- A thorough validation process was conducted to identify and remove any corrupted or low-quality images, ensuring data integrity and reliability;
- Images were resized to 224×224 pixels to maintain consistency;
- Pixel values were normalized to the range [0,1] to enhance numerical stability;
- The datasets were structured using an 80-20 training-validation split to ensure a balanced and robust training process.





Fig. 1. Samples of mineral specimens used in this study. Each mineral is shown with its name and chemical formula.

3.2. Training Phase

We employed the Vision Transformer (ViT) backbone, which is optimized for extracting high-dimensional visual features from input images. The BLIP model leverages a bootstrapping approach, progressively refining its learning by iteratively improving visual-textual alignment. enhancing its generalization across diverse mineral images. The training process can be broken down into the following steps:

1. Feature Extraction: Each mineral image from both datasets is fed into the BLIP model, where

the ViT backbone extracts visual features such as texture, color, and patterns.

- 2. Caption Generation: The extracted visual features are processed by BLIP's text decoder to generate descriptive captions, providing contextual information about minerals, including color variations, surface textures, and structural properties.
- 3. Noise Filtering: BLIP's unique filtering mechanism compares the generated captions with potential noise or irrelevant information, to retain only meaningful and relevant textual descriptions, enhancing the model's ability to

focus on discriminative features and improve classification accuracy.

4. Adaptation for Classification: To adapt BLIP for classification instead of captioning, the final laver of the model was modified. Specifically, the text decoder was replaced with a classification head, which maps the visual-textual embeddings to the respective mineral classes. This modification allows the to perform classification model while maintaining the benefits of BLIP's multimodal approach.

3.3. Embedding Extraction and Classification

To further refine the classification process, we extracted visual embeddings using the BLIP processor to process input images and extract visual information. These embeddings were then used as input for a classifier, which is trained to predict the mineral categories.

To evaluate the model's performance, we employed the following metrics:

- Accuracy;
- Precision;
- Recall;
- F1 Score;
- Matthews Correlation Coefficient (MCC);
- Youden's Index (J-Index).

4. Datasets and Experimental Results

In this section, we first describe the datasets used to evaluate the proposed MineralBLIP model and the baseline CNN. We then present and discuss the experimental findings.

4.1. Datasets

Two datasets were used in our experiments:

- Dataset 1: Comprising 6172 images across four mineral classes. This dataset was split 80/20 for training and validation, with 1233 images reserved for validation;
- Dataset 2: Containing 2310 images representing five mineral classes, also divided into training and validation sets, with 459 images used for validation.

These datasets provide a range of mineral images, ensuring diversity and sufficient complexity to evaluate the robustness of both models.

4.2. Experimental Results

A comprehensive performance comparison between MineralBLIP and the baseline CNN is presented in Table 1 and visually illustrated in Fig. 2a and Fig. 2b. As shown, MineralBLIP achieves consistent improvements over the CNN model on both datasets:

- Dataset 1: MineralBLIP attains an accuracy of 86 %, precision of 86 %, recall of 86 %, and F1-score of 86 %. In contrast, the CNN model records 76 % accuracy, 77 % precision, 76 % recall, and 76 % F1-score;
- Dataset 2: MineralBLIP achieves an accuracy of 84 %, 87 % precision, 81 % recall, and an F1-score of 82 %, surpassing the CNN's 76 % accuracy, 77 % precision, 76 % recall, and 74 % F1-score.

Table 1. Perfo	rmance comparis	son between t	he CNN-based
r	nodel and Miner	alBLIP mode	l.

		MineralBLIP	CNN
Dataset 1	Accuracy	0.86	0.76
	Precision	0.86	0.77
	Recall	0.86	0.76
	F1 Score	0.86	0.76
	MCC	0.82	0.69
	J-Index	0.76	0.62
		MineralBLIP	CNN
Dataset 2	Accuracy	0.84	0.76
	Precision	0.87	0.77
	Recall	0.81	0.76
	F1 Score	0.82	0.74
	MCC	0.80	0.70
	J-Index	0.71	0.61

4.3. Discussion

A detailed examination of these results reveals that MineralBLIP effectively addresses key challenges in fine-grained mineral classification, such as intra-class variability and inter-class similarity. CNNs, which rely heavily on edge and texture cues, often misclassify minerals with closely resembling color distributions. In contrast, MineralBLIP's multi-head self-attention mechanism enhances feature extraction, enabling the model to focus on the most discriminative regions within each image. This capability leads to superior differentiation among visually similar mineral categories.

These findings underscore the advantages of transformer-based vision models similar to MineralBLIP for complex classification tasks, particularly when precise discrimination is essential.

4.4. Computational Efficiency

The computational efficiency of MineralBLIP was evaluated against the CNN model on both datasets:

• Dataset 1: The CNN model required 19 minutes and 53 seconds for inference, while MineralBLIP completed inference in 5 minutes and 7 seconds;
• Dataset 2: The CNN model took 5 minutes and 53 seconds for inference, whereas MineralBLIP finished in 1 minute and 52 seconds.



Fig. 2a. MineralBLIP model and baseline CNN model performance on Dataset 1.



Fig. 2b. MineralBLIP model and baseline CNN model performance on Dataset 2.

These results highlight the improved computational efficiency of MineralBLIP.

5. Conclusion

In this study, we introduce MineralBLIP, a vision-language model based on Vision Transformers for fine-grained mineral classification. Utilizing two datasets, one comprising 6172 images spanning four mineral classes and another consisting of 2310 images representing five mineral classes. Our experiments demonstrate that conventional CNNs, despite their effectiveness in visual data classification, require high-quality, diverse datasets. In contrast, MineralBLIP leverages both visual and textual modalities through vision-language pre-training to achieve substantial performance gains. Specifically, MineralBLIP attained an average F1-score of 84 %, outperforming the baseline CNN model's 75 %. These findings underscore the efficacy of transformer-based architectures in capturing subtle features with minimal customization or fine-tuning, thereby enhancing the efficiency of pre-training on large-scale datasets.

Our key contributions are summarized as follows:

• Novel Approach: We introduce MineralBLIP, a vision-language model that integrates visual and textual modalities to overcome the limitations of

conventional CNNs in fine-grained mineral classification;

- Enhanced Performance: Extensive experiments on two mineral image datasets demonstrated that MineralBLIP significantly outperforms CNN-based models across multiple key metrics;
- Methodological Insights: Our analysis provided valuable insights into how vision-language pre-training and multi-head self-attention mechanisms improve feature extraction and class differentiation in complex imaging scenarios.

References

- N. Alqahtani, R. T. Armstrong, P. Mostaghimi, Deep learning convolutional neural networks to predict porous media properties, in *Proceedings of the SPE Asia Pacific Oil and Gas Conference and Exhibition*, Australia, 23-25 October 2018, SPE-191906-MS.
- [2]. Y. LeCun, Y. Bengio, G. Hinton, Deep learning, *Nature*, Vol. 521, 2015, pp. 436-444.
- [3]. K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, United States, 27-30 June 2016, pp. 770-778.
- [4]. E. E. Baraboshkin, L. S. Ismailova, D. M. Orlov, E. A. Zhukovskaya, G. A. Kalmykov, O. V. Khotylev, et al., Deep convolutions for in-depth automated rock typing, *Computers & Geosciences*, Vol. 135, 2020, 104330.
- [5]. B. Wu, D. Meng, L. Wang, N. Liu, Y. Wang, Seismic impedance inversion using fully convolutional residual network and transfer learning, *IEEE Geoscience and Remote Sensing Letters*, Vol. 17, Issue 12, 2020, pp. 2140-2144.
- [6]. R. P. de Lima, F. Suriamin, K. J. Marfurt, M. J. Pranter, Convolutional neural networks as aid in core lithofacies classification, *Interpretation*, Vol. 7, Issue 3, 2019, pp. SF27-SF40.
- [7]. A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, et al., An image is worth 16x16 words: Transformers for image recognition at scale, *arXiv preprint*, 2020, arXiv:2010.11929.
- [8]. J. Li, D. Li, C. Xiong, S. Hoi, Blip: Bootstrapping language-image pre-training for unified vision-language understanding and generation, in *Proceedings of the 39th International Conference on Machine Learning*, United States, 17-23 July 2022, pp. 12888-12900.
- [9]. Y. Bazi, L. Bashmal, M. M. A. Rahhal, R. A. Dayil, N. A. Ajlan, Vision transformers for remote sensing image classification, *Remote Sensing*, Vol. 13, Issue 3, 2021, 516.
- [10]. J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, in *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Vol. 1, United States, 2-7 June 2019, pp. 4171-4186.
- [11]. Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, et al., Swin transformer: Hierarchical vision transformer

using shifted windows, in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Canada, 11-17 Oct 2021, pp. 10012-10022.

- [12]. A. Krizhevsky, I. Sutskever, G. E. Hinton, ImageNet classification with deep convolutional neural networks, in Advances in Neural Information Processing Systems, Vol. 25, *Curran Associates, Inc.*, 2012.
- [13]. J. Maurício, I. Domingues, J. Bernardino, Comparing vision transformers and convolutional neural networks for image classification: a literature review, *Applied Sciences*, Vol. 13, Issue 9, 2023, 5521.
- [14]. Y. Attallah, E. Zigh, A. P. Adda, Optimized 3D-2D CNN for automatic mineral classification in hyperspectral images, *Reports on Geodesy and Geoinformatics*, Vol. 118, Issue 1, 2024, pp. 82-91.
- [15]. E. Brempong, M. Agangiba, D. Aikins, MiNet: A convolutional neural network for identifying and categorising minerals, *Ghana Journal of Technology*, Vol. 5, Issue 1, 2020, pp.86-92
- [16]. J. I. Cifuentes, L. E. Arias, E. Pirard, F. Castillo, Mineral classification using convolutional neural networks and SWIR hyperspectral imaging, *Proceedings of SPIE*, Vol. 12903, 13 March 2024, 1290309.
- [17]. L. Xu, J. Xie, F. Cai, J. Wu, Spectral classification based on deep learning algorithms, *Electronics*, Vol. 10, Issue 16, 2021, 1892.
- [18]. X. Liu, C. Aldrich, Multivariate image processing in minerals engineering with vision transformers, *Minerals Engineering*, Vol. 208, 2024, 108599.
- [19]. M. He, M. Yang, Z. Zhang, Y. Chen, Y. Wang, X. Zheng, Mineral identification based on multi-label image classification, *Minerals*, Vol. 12, Issue 11, 2022, 1338.
- [20]. L. Liu, J. E. Santos, M. Prodanović, M. J. Pyrcz, Mitigation of spatial nonstationarity with vision

transformers, Computers & Geosciences, Vol. 178, 2023, 105412.

- [21]. A. Koeshidayatullah, S. Al-Azani, E. Baraboshkin, M. Alfarraj, FaciesViT: vision transformer for an improved core lithofacies prediction, *Frontiers in Earth Science*, Vol. 10, 2022, 992442.
- [22]. Q. Chen, Y. Hong, MedBLIP: bootstrapping languageimage pre-training from 3D medical images and texts, in *Proceedings of 17th Asian Conference on Computer Vision*, Vietnam, 8-12 December 2024, pp. 98-113.
- [23]. Y. Xiao, Y. Dou, S. Yang, PointBLIP: a zero-training point cloud classification network based on BLIP-2, *Remote Sensing*, Vol. 16, Issue 13, 2024, 2453.
- [24]. L. Tao, H. Zhang, H. Jing, Y. Liu, D. Yan, G. Wei, X. Xue, Advancements in vision-language models for remote sensing: datasets, capabilities, and enhancement techniques, *Remote Sensing*, Vol. 17, Issue 1, 2024, 162.
- [25]. H. Nguyen, T. A. Nguyen, Hybrid vision transformers and CNNs for enhanced transmission line segmentation in aerial images, *International Journal of Advanced Computer Science and Applications (IJACSA)*, Vol. 15, Issue 1, 2024, pp. 434-442.
- [26]. X. Liu, K. Hu, S. E. Grasby, B. Lee, A hybrid machine learning model for improving regression of mineral composition estimation using well logging data, in *Proceedings of the Fourth International Meeting for Applied Geoscience & Energy (IMAGE'24)*, United States, 26-29 August, pp. 449-453.
- [27]. A Mineral Database, https://www.mindat.org/
- [28]. MultiLabel Classification Dataset, Roboflow Universe, https://universe.roboflow.com/finalproject-flqsu/ multilabel-classification-vbxvr

(066)

An Improved Algorithm for Computing Matroids over Polynomials

David W. Ash

Real Time Agents Inc, 191 Twinbridge Cir, Pleasant Hill, CA 94523-4739, USA Tel.: +1 650 847 9893 E-mail: dash@alumni.stanford.edu

Summary: Artificial intelligence models sometimes require the computation of matroid. A matroid is a set of subsets of a given finite set considered to be the independent subsets of that set. Matroids sometimes arise in biological models of the human body [1], where it can be important to know whether a set of species is considered independent or not. However such species are often represented by complex polynomial and other functions and determining linear independence, which is necessary to compute the matroid, can be computationally expensive. We propose a way of decreasing the computational complexity of computing matroids which will be useful for computation of matroids in biological and other domains.

Keywords: Matroids, Bases, Circuits, Biological models, Computational complexity, Polynomial dependence.

1. Introduction

Biological models, such as the Wnt signaling pathway, have important applications in understanding development, adult stem tissue maintenance, and many diseases including cancer. Therefore, gaining an in depth mathematical understanding of these models can be useful in determining which groups of variables are related in different biological situations. The BioModels website [2] contains over 1000 manually curated models of biological processes. One of these is the Wnt signaling pathway model developed by Lee ([3, 4]). MacLean et.al. [1] have proposed the use of matroids as a mathematical tool for analyzing such models. Matroids are a useful tool because they allow parameter free analysis of models. In addition to their use for biological models, matroids have also found application in signal processing [7]. A matroid is a mathematical structure, derived from linear algebra but more abstract, containing sets of dependent and independent variables. The computation of a matroid for any given model can be accomplished using the Gröbner basis approach recommended by MacLean et.al. [5] for simpler models. However computational complexity issues render the Gröbner basis approach impractical for more complex models. Therefore MacLean et.al. recommend using linearization to compute matroids.

2. Linearization

The linearization approach, along with an introduction to matroids, was originally proposed by Ingleton [6]. A matroid is a finite set on which certain subsets, equivalent to maximal linearly independent subsets, are referred to as bases. Other subsets, equivalent to minimal linearly dependent subsets, are referred to as circuits. In the context of biological models such as the Wnt signaling pathway model, a basis would be a maximal set of species algebraically independent over their underlying parameters.

Following Ingleton, suppose that the set of species is given by $s_1, s_2, ..., s_m$, where the s_i are functions of a set of parameters $x_1, x_2, ..., x_r$. We can define a gradient vector ∇s_i as $\nabla s_i = (D_1 s_i, D_2 s_i, ..., D_r s_i)$ where $D_j s_i$ is the symbolic partial derivative of s_i with respect to x_j . Then Ingleton's theorem [6] tells us that the species $s_1, s_2, ..., s_m$ are algebraically independent over parameters $x_1, x_2, ..., x_r$ iff the gradients $\nabla s_1, \nabla s_2, ..., \nabla s_m$ are linearly independent.

So, for example, given the details of the Wnt signaling pathway model provided in [5], the species are given by D_i , D_a , Y_a , Y_i , G, C_{NA} , A, C_{XY} , C_{XYp} , X_p , X, N, T, C_{XT} and C_{XA} . The parameters are given by α_1 , $\alpha_2, \ldots, \alpha_{22}$. We can therefore provide a couple of the gradients. For example, $\nabla D_i = (-D_i, D_a, 0, ..., 0)$ and $\nabla D_a = (D_i, -D_a, 0, \dots, 0)$. As these two gradients are linearly dependent, it follows that the two species D_i and D_a are algebraically dependent. As they form a minimal algebraically dependent set, it therefore follows that species D_i and D_a form a circuit in the associated matroid. This provides us with a mechanism for computing bases and circuits. However, as this computation requires symbolic manipulations over a large range of algebraic symbols, it is not that computationally efficient.

3. Integerization

To address this lack of computational efficiency, we propose an approach we will call integerization. To perform integerization, we construct an integer mapping function, *P*, that maps each s_i and each x_j to a unique prime number. For example, we might set $P(D_i) = 11$, $P(D_a) = 13$, ..., $P(C_{XA}) = 67$, $P(\alpha_1) = 71$, ..., $P(\alpha_{22}) = 179$ in the Wnt signaling pathway model. We can then extend the definition of *P* to the ring of polynomials with integer coefficients $\mathbb{Z}[s_1, s_2, ..., s_m; x_1, x_2, ..., x_r] = \mathbb{Z}[s; x]$ as follows:

• For $k \in \mathbb{Z}$, $p \in \mathbb{Z}[s; x]$, set P(kp(s; x)) = kP(p(s; x));

- For $p, q \in \mathbb{Z}[s; x]$, set $P(p(s; x) \cdot q(s; x)) = P(p(s; x)) \cdot P(q(s; x));$
- For $p, q \in \mathbb{Z}[s; x]$, set P(p(s; x) + q(s; x)) = P(p(s; x)) + P(q(s; x)).

We can finally perform a similar mapping on the gradient vector to get $P(\nabla s_i) = (P(D_1s_i), P(D_2s_i), \dots, P(D_rs_i))$. This reduces the gradient vector to a vector over the integers. We can then show the following results:

Theorem. The following results hold:

- If a set of species s_{i1}, s_{i2}, ..., s_{ik} is algebraically dependent over parameters x₁, x₂, ..., x_r, then the corresponding integer gradients P(∇s_{i1}), P(∇s_{i2}), ..., P(∇s_{ik}) are linearly dependent;
- If, for a set of species, the corresponding integer gradients $P(\nabla s_{i_1})$, $P(\nabla s_{i_2})$, ..., $P(\nabla s_{i_k})$ are linearly independent, then the set of species $s_{i_1}, s_{i_2}, \ldots, s_{i_k}$ is algebraically independent over parameters x_1, x_2, \ldots, x_r .

Given this theorem, the following revised algorithm for computing bases over a matroid is indicated:

Algorithm. Find candidate bases by looking for linearly independent subsets of the set of integer gradients. When such a set is found:

- Step 1: If it is the first candidate basis, we know that it may correspond to an algebraically independent set of species.
 - **Step 1.1**: Use traditional algebraic techniques to verify its maximality.
- Step 2: If it is a subsequent candidate, determine if it has the same cardinality as previously found bases. If the cardinality matches it is a basis.

This algorithm provides for the possibility of a significant speedup in calculating bases because the more computationally expensive algebraic calculations only need to be done once. Indeed, empirically we have found that the verification step can, in practice, be eliminated. Although it is not theoretically guaranteed that a maximally linearly independent set of species in the integer space will correspond to a maximal linearly independent set of species in the algebraic space, we have found that in practice this is always the case.

4. Results

The integerization approach described in Section 3 was compared to the existing Gröbner basis approach discussed in Section 1. Both approaches were run specifically on the Lee et.al. model described in Section 1.1 of MacLean et.al. [5]. The results showed that using the linearization approach took 1.5 seconds to compute the circuits and bases of the matroid, whereas using the Gröbner basis approach took 179 seconds to perform the same computation – a speedup of a factor of about 119.

However, in doing so we have skipped Step 1.1 in the algorithm described in the previous section, and have simply assumed that the candidate basis will always be maximal. Given that the prime numbers associated with parameters and species are essentially assigned randomly, there is some risk that by skipping this step, we will (at random) pass a candidate basis which is not, in fact, maximal. It has been found that this is, in practice, rarely the case. This was tested on the shuttle model described in Section 2 of MacLean et.al. [5]. An experiment with 100 trials was run, with the prime numbers being assigned randomly each time. The results were the same each time, so in practice for every trial, skipping Step 1.1 did not result in any change to the results. However, this is for now merely an empirical observation.

5. Conclusions and Next Steps

Based on the results from Section 4, we can conclude - very tentatively - that the integerization approach is a substantial improvement on existing Gröbner basis approach for computing matroids. However additional work is definitely needed. The main point that is missing so far are results where we use linearization, but without proceeding to integerization. It would be useful to compare the computational efficiency of linearization without integerization (Section 2) with that of integerization (Section 3). If linearization without integerization turns out to be nearly as good as integerization, it would speak against proceeding to integerization. The conjecture would be that linearization without integerization is much more computationally expensive, but this needs to be verified.

Moreover, results where we include Step 1.1 of the algorithm – which are currently omitted – should also be determined. We need to determine the computational complexity of adding Step 1.1 to the algorithm. We also need to determine the likelihood – is it vanishingly unlikely or a real possibility – of omitting Step 1.1 causing an error in the computation of the matroid.

Most of the results so far are empirical. As a further next step, theoretical estimates of the computational complexity of the various variants of the algorithm should also be determined. Additionally, a theoretical estimate of the probability that Step 1.1 turns out to be needed is necessary.

References

- [1]. A. L. MacLean, Z. Rosen, H. M. Byrne, H. A. Harrington, Parameter-free methods distinguish Wnt pathway models and guide design of experiments, *Proceedings of the National Academy of Sciences*, Vol. 112, Issue 9, 2015, pp. 2652-2657.
- [2]. European Molecular Biology Laboratory, European Bioinformatics Institute, BioModels, 2006, https://www.ebi.ac.uk/biomodels
- [3]. E. Lee, A. Salic, R. Krüger, R. Heinrich, M. W. Kirschner, The roles of APC and Axin derived from experimental and theoretical analysis of the WNT pathway, *PLoS Biology*, Vol. 1, Issue 1, 2003, E10.

7th International Conference on Advances in Signal Processing and Artificial Intelligence (ASPAI' 2025), 8-10 April 2025, Innsbruck, Austria

- [4]. V. K. Kothamachu, M. G. Roberts, Lee2003 Roles of APC and Axin in Wnt Pathway (Without Regulatory Loop), at European Molecular Biology Laboratory, European Bioinformatics Institute, BioModels, https://www.ebi.ac.uk/biomodels/services/download/g et-files/MODEL1708310000/4/ BIOMD0000000658.pdf
- [5]. A. L. MacLean, Z. Rosen, H. M. Byrne, H. A. Harrington, Supporting Information for: Parameter-Free Methods Distinguish WNT Pathway Models and Guide Design of Experiments,

https://pmc.ncbi.nlm.nih.gov/articles/instance/435282 7/bin/pnas.1416655112.sapp.pdf

- [6]. A. W. Ingleton, Representation of matroids, in Combinatorial Mathematics and its Applications (D. J. A. Welsh, Ed.), *Academic Press*, London, New York, 1971, pp. 149-167.
- [7]. E. Tohidi, R. Amiri, M. Coutino, D. Gesbert, G. Leus, A. Karbasi, Submodularity in action: from machine learning to signal processing applications, *IEEE Signal Processing Magazine*, Vol. 37, Issue 5, 2020, pp. 120-133.

